

الإحصاء

لِلإِدَارِيِّينَ وَالْإِقْتِصَادِيِّينَ

د. دلال القاضي

أستاذ مشارك

جامعة بغداد / جامعة عمان الأهلية

د. سهيلة عبد الله

أستاذ مشارك

جامعة بغداد / جامعة عمان الأهلية

د. محمود البياتي

أستاذ مشارك / تحليل بيانات

جامعة بغداد / جامعة عمان الأهلية



الإحصاء للإداريين والإقتصاديين

د. سهيلة عبدالله

أستاذ مشارك

جامعة بغداد/جامعة عمان الأهلية

د. دلال القاضي

أستاذ مشارك

جامعة بغداد/جامعة عمان الأهلية

د. محمود البياتي

أستاذ مشارك / تحليل بيانات

جامعة بغداد/جامعة عمان الأهلية

2005م



محفوظ جميع الحقوق

رقم الإيداع لدى دائرة المكتبة الوطنية (٢٠٠٣/٧/١٤٦٨)

٥١٩.٥٣

القاضي، دلال

الإحصاء للإداريين والإقتصاديين / دلال القاضي، سهيلة عبدالله، محمود البياتي.

عمان: دار ومكتبة الحامد، ٢٠٠٣.

(٣٧٦) ص

ر.إ. (٢٠٠٣/٧/١٤٦٨)

الواصفات: / الإحصاء الوصفي /

* تم إعداد بيانات الفهرسة والتصنيف الأولية من قبل دائرة المكتبة الوطنية

* رقم الإجازة المتسلسل لدى دائرة المطبوعات والنشر ٢٠٠٣/٧/١٥٣٢

* (ردمك) ISBN 9957-32-039-4



دار الحامد للنشر والتوزيع

الأردن - عمان

هاتف: ٥٢٣١٠٨١ فاكس: ٥٢٣٥٥٩٤ - ٥٠٩٦٢٦

ص.ب.: ٣٦٦ عمان ١١٩٤١ الأردن

E-mail: dar_alhamed@hotmail.com

لا يجوز نشر أو اقتباس أي جزء من هذا الكتاب، أو اختزان مادته بطريقة الاسترجاع، أو نقله على أي وجه، أو بأي طريقة أكانت إلكترونية، أم ميكانيكية، أم بالتصوير، أم التسجيل، أم بخلاف ذلك، دون الحصول على إذن الناشر الخطي، وبخلاف ذلك يتعرض الفاعل للملاحقة القانونية

التصميم والإخراج الفني: لينو ابراهيم

الفصل الأول

المقدمة ووصف البيانات

- 1-1 مقدمة التعريف بعلم الإحصاء 15
- 1-2 طبيعة البيانات 20
- 1-3 جمع البيانات 22
- 1-4 عرض البيانات الإحصائية 23
- 1-4-1 العرض الجدولي والتوزيع التكراري 23
- 1-4-2 التوزيع التكراري النسبي 31
- 1-4-3 التوزيع التكراري المتجمع 32
- 1-5 التمثيل البياني للتوزيعات التكرارية 35
- 1-6 أشكال التوزيعات التكرارية 42

الفصل الثاني

مقاييس النزعة المركزية والتشتت

- 2-1 مقدمة 53
- 2-2 الوسط الحسابي 53
- 2-2-1 الوسط الحسابي للبيانات الخام 54
- 2-2-2 الوسط الحسابي للبيانات المجمعة 54
- 2-2-3 خواص الوسط الحسابي 56
- 2-3 الوسيط 64
- 2-3-1 الوسيط للبيانات الخام 65
- 2-3-2 الوسيط للبيانات المجمعة 66

69	2-4 المنوال
69	2-4-1 المنوال للبيانات الخام
69	2-4-2 المنوال للبيانات المجمعة
72	2-5 الوسط الهندسي
72	2-5-1 الوسط الهندسي للبيانات الخام
73	2-5-2 الوسط الهندسي للبيانات المجمعة
74	2-6 الربيعيات والعشيرات (المئينات) والمدى الربيعي
78	2-7 العلاقة بين مقاييس النزعة المركزية (الوسط الحسابي والوسيط والمنوال)
81	2-8 المدى
82	2-9 الانحراف المتوسط
82	2-9-1 الانحراف المتوسط للبيانات الخام
84	2-9-2 الانحراف المتوسط للبيانات المجمعة
85	2-10 التباين والانحراف المعياري
85	2-10-1 التباين والانحراف المعياري للبيانات الخام
88	2-10-2 التباين والانحراف المعياري للبيانات المجمعة
90	2-10-3 تفسير الانحراف المعياري
92	2-11 معامل الاختلاف أو التغير
93	2-12 الدرجة المعيارية
95	2-13 الغصن والورقة
99	2-14 الرسم الصندوقي
100	2-14-1 كيفية بناء الرسم الصندوقي البسيط
102	2-14-2 كيفية بناء الرسم الصندوقي المعدل
103	2-14-3 استخدامات أخرى للرسم الصندوقية

الفصل الثالث

الانحدار والارتباط الخطي البسيط

- 111 3-1 مقدمة
- 112 3-2 معامل الارتباط الخطي البسيط
- 115 3-3 معامل الارتباط للرتب والصفات
- 117 3-4 الانحدار الخطي البسيط
- 129 3-5 العلاقة بين معاملات الانحدار ومعامل الارتباط

الفصل الرابع

الاحتمال والمنتزيع العشوائي

- 135 4-1 مقدمة
- 135 4-2 نظرية المجموعات
- 142 4-3 التجربة ، فضاء العينة والحدث
- 145 4-4 الاحتمال ومعناه
- 147 4-4-1 قوانين الاحتمالات
- 151 4-4-2 طرق العد
- 160 4-4-3 الاحتمال الشرطي
- 162 4-4-4 الأحداث المستقلة
- 166 4-4-5 نظرية بيز
- 168 4-5 المتغيرات العشوائية والتوزيعات الاحتمالية
- 169 4-5-1 التوزيع الاحتمالي للمتغير العشوائي المنفصل
- 173 4-5-2 التوزيع الاحتمالي للمتغير العشوائي المستمر
- 175 4-5-3 القيمة المتوقعة (التوقع الرياضي)

181	4-6 أمثلة التوزيعات المتقطعة
181	4-6-1 توزيع ذي الحدين
185	4-6-2 توزيع بواسون
186	4-6-3 التوزيع فوق الهندسي
188	4-6-4 التوزيع الهندسي
188	4-6-5 التوزيع المنتظم
190	4-7 أمثلة التوزيعات المستمرة
190	4-7-1 التوزيع الطبيعي
192	4-7-2 التوزيع الطبيعي المعياري
196	4-7-3 التوزيع المنتظم
197	4-7-4 توزيع جاما
198	4-7-5 توزيع بيتا

الفصل الخامس

توزيعات المعاينة

207	5-1 نظرية المعاينة
208	5-2 أنواع العينات
208	5-2-1 العينة العشوائية البسيطة
209	5-2-2 العينة المنتظمة
209	5-2-3 العينة العشوائية الطبقية
211	5-2-4 العينة العشوائية المتعددة المراحل
211	5-3 توزيع المعاينة
212	5-3-1 توزيع المعاينة للوسط الحسابي من مجتمع طبيعي
214	5-3-2 توزيع المعاينة للفرق بين وسطين
216	5-3-3 توزيع المعاينة للنسب

- 218 5-3-4 توزيع المعاينة للفرق ما بين نسبتين
- 219 5-3-5 توزيع المعاينة للتباين

الفصل السادس

التقدير واختبار الفرضيات

- 225 6-1 نظرية التقدير
- 225 6-1-1 التقدير النقطي
- 226 6-1-2 التقدير بفترة
- 245 6-2 اختبار الفرضيات
- 250 6-2-1 اختبارات تتعلق بالمتوسطات: حجم العينة كبير
- 261 6-2-2 اختبارات تتعلق بالمتوسطات: حجم العينة صغير
- 270 6-2-3 اختبارات تتعلق بالنسب
- 277 6-3 استخدام طريقة P-Value لاختبار الفرضيات

الفصل السابع

تحليل التباين

- 279 7-1 مقدمة
- 290 7-2 اختبار t-test
- 292 7-3 تحليل التباين
- 299 7-4 المقارنات المتعددة
- 301 7-5 اختبار تساوي عدة تباينات

الفصل الثامن الطرق الالامصلمية

- 311 8-1 مقدمة
- 311 8-2 اختبار "ولكوكسن"
- 316 8-3 اختبار "مان - ويتني"
- 320 8-4 اختبار "ولكوكسن" للفرق المزدوج
- 322 8-5 اختبار "كروسكل - والس"
- 325 8-6 اختبار معامل الارتباط
- 327 8-7 اختبار حسن المطابقة
- 330 8-8 اختبار الاستقلالية

الفصل التاسع الأرقام القياسية والسلاسل الزمنية

- 339 9-1 مقدمة
- 339 9-2 السلسلة الزمنية والأرقام القياسية
- 340 9-2-1 الرقم القياسي البسيط
- 342 9-2-2 الرقم القياسي التجميعي البسيط
- 343 9-2-3 الرقم القياسي للاسبير
- 343 9-2-4 الرقم القياسي لباش
- 344 9-2-5 الرقم القياسي الأمثل لفشر
- 345 9-2-6 الرقم القياسي الحقيقي للدخل
- 347 9-3 تحليل السلاسل الزمنية
- 347 9-3-1 طريقة الانحدار الخطي البسيط
- 349 9-3-2 طريقة المتوسطات المتحركة
- 353 9-3-3 طريقة التنعيم
- 359 ملحق بالجدول الاحصائية
- 375 المراجع

بعون الله فهذه الطبعة الثانية من كتاب "الإحصاء للإداريين والاقتصاديين" والذي نالت الطبعة الأولى منه حضوراً طيباً واستخدماً واسعاً من قبل الطلبة وكثير من الباحثين والأساتذة. واتسمت الطبعة الثانية هذه بأنها تعتبر نسخة منقحة من الطبعة الأولى للأسباب التالية:

1- تم اكتشاف وتلافي وتصحيح كافة الأخطاء المطبعية واللغوية والفنية التي كانت موجودة .

2- تم إضافة مباحث عديدة يمكن اعتبارها جزءاً مهماً يزيد من متانة المادة العلمية الموجودة في الطبعة الأولى . منها مبحث لتحديد ورصد أشكال التوزيعات التكرارية في الفصل الأول وإضافة مبحث للعلاقات بين مقاييس النزعة المركزية وآخر عن تفسير الانحراف المعياري في الفصل الثاني . كما تم إضافة مبحث للعلاقة بين معامل الارتباط ومعاملات الانحدار في الفصل الثالث .

3- تم إضافة العديد من الأمثلة في الفصول المختلفة ، وكذلك إضافة العديد من الأسئلة في نهاية كل فصل .

4- كما تم إضافة التطبيق الإحصائي على الحاسب الإلكتروني ، موضحين فيه كافة الخطوات والأوامر اللازمة لتنفيذ الطرق عن طريق الضغط على الأزرار المعينة وباستخدام البرامج الجاهزة SPSS وكذلك EXCEL . ويمكن الاتصال على العنوان التالي حول أي سؤال على الجانب التطبيقي إلى أنظمة الحاسوب :
m_bayati@hotmail.com

والله ولي التوفيق

المؤلفين

1

الفصل الأول

مقدمة ووصف البيانات

Introduction and Data Description

1-1 مقدمة التعريف بعلم الإحصاء

1-2 طبيعة البيانات

1-3 جمع البيانات

1-4 عرض البيانات الإحصائية

1-4-1 العرض الجدولي

1-4-2 التوزيع التكراري النسبي

1-4-3 التوزيع التكراري المتجمع

1-5 التمثيل البياني للتوزيعات التكرارية

1-6 أشكال التوزيعات التكرارية

الجزء
الاول

1

الفصل الأول

المقدمة ووصف البيانات

Introduction and Data Description

1-1 مقدمة للتعريف بعلم الإحصاء:

عرف الإحصاء قديماً وتم استخدامه من قبل الفراعنة في بناء الأهرامات حيث قاموا بتعداد لسكان مصر وثروتها واستخدموا النتائج في تنظيم مشروع البناء. وكذلك في عصر الدولة الإسلامية استخدم الخليفة المأمون فكرة الحصر والعد لمعرفة عدد السكان ومقدار الزكاة وكان استخدام الإحصاء في البداية مقصوراً على الأعمال الخاصة بشؤون الدولة كما يدل على ذلك الأصل اللغوي في اسم هذا العلم وهو "Statistics" حيث كلمة "State" تعني الدولة.

أما معنى إحصاء لأي فرد فكان مقتصرًا على الجداول العددية التي تصف ظاهرة معينة أو على الرسوم البيانية أو الأشكال التصويرية التي تعرض التغيرات في ظاهرة خلال فترة زمنية معينة. ويمكن ملاحظة ذلك من خلال حياتنا اليومية مثلاً التطلع في الصحف اليومية حيث يمكن أن تشاهد بعض الجداول التي تبين معدل كميات نزول المطر أو تمثيل بيانات عن أسواق النقد والتقلبات في العملة والأسهم... الخ من المعلومات والرسومات الإحصائية. أما اليوم فقد أصبح للإحصاء أهمية كبيرة في كثير من المفردات اليومية.

ومن هذا يمكن إعطاء وصف صريح للإحصاء:

فهو علم يهتم بجمع البيانات والملاحظات ومن ثم يتم تنظيمها وعرضها باستخدام الأساليب العلمية لتحليلها واستخلاص النتائج منها، ولعلم الإحصاء علاقة وطيدة بمختلف العلوم الأخرى منها الرياضيات، العلوم الإنسانية، علم الاجتماع والدراسات السكانية، وكذلك العلوم الطبية والهندسية وعلم الأرض والحياة والوراثة وغيرها من العلوم.

ويعتمد الباحثين بشكل واسع على دراسة العينات ليحصل على البيانات الإحصائية التي تتعلق بنشاطات الإنسان والأحوال المتعلقة به، وإن كثير من الطرق الإحصائية قد تم استعمالها وتستخدم في الوقت الحاضر لجمع وتحليل وعرض البيانات لغرض التخطيط الاقتصادي واتخاذ القرارات لذا يلعب التحليل الإحصائي دورا بارزا في كثير من حقول النشاط الإنساني والذي يعتبر مفيدا جدا في تبادل المعلومات والوصول إلى الاستنتاجات والاستدلالات في البيانات ومن ثم في الإرشاد إلى التخطيط المنطقي واتخاذ القرار.

ويمكن تقسيم الإحصاء إلى قسمين رئيسيين:

1- الإحصاء الوصفي Descriptive statistics:

و الذي يمثل الطرق الرقمية أو الحسابية لجمع المعلومات والبيانات لتلخيصها واختصارها ومن ثم عرض المعلومات عن طريق الجداول والرسوم البيانية وغيرها.

2- الإحصاء الاستدلالي Inferential statistics:

وهو ذلك الجزء من الإحصاء الذي يعنى بتحليل البيانات المتوفرة في العينة Sample كأساس لتحليل البيانات الموجودة في المجتمع Population للتوصل إلى أساليب التقدير والاختبار واتخاذ القرارات والتنبؤ أو الاستقراء، ويلعب هذا الجزء دورا مهما في تخطيط التجارب التي تجمع منها البيانات ومن ثم تصميمها. لذا فإن الإحصاء يهتم بطرق جمع البيانات وتمثيلها وعرضها "الإحصاء الوصفي" ومن ثم تحليلها وتفسيرها والتوصل إلى الاستنتاجات "الإحصاء الاستدلالي".

مراحل البحث العلمي:

يعتمد البحث العلمي على الطرق والأدوات الإحصائية المختلفة والمستخدم في أي من العلوم أو أي مجال حسب طبيعة ونوع البحث، ومن أهم مراحل البحث العلمي هي:

1- تحديد المشكلة:

يتم تحديد نوع المشكلة وفي أي مجال أو في أي علم من العلوم التي ذكرت سابقا التي تستحق البحث و التقصي فدور الباحث هنا كيف يتم اختيار المشكلة المناسبة لدراسة الظواهر الغريبة أو المألوفة، وذلك بالتعبير عنها بعلاقة ما بين المتغيرات Variables أو المشاهدات Observations بحيث تمكن الباحث من إجراء التحليل الإحصائي أو الاستنتاجي.

2- استخدام الأساليب الإحصائية:

بعد تحديد المشكلة للدراسة والإلمام بجميع جوانبها يتم تفسير الظواهر تفسيراً علمياً من جميع الجوانب وتحديد الطرق التي سوف تستخدم لحل المشكلة ليتمكن الباحث من اختيار البيانات الملائمة لها وفي أي جانب التوجه والتي يتم فيها مرحلة جمع البيانات.

3- جمع المعلومات أو البيانات:

تعتمد على بعض الأساليب الإحصائية في جمع المعلومات " البيانات " والتي سوف نتطرق لها لاحقاً بشكل مفصل.

4- تحليل البيانات:

استخدام الطرق والأساليب الإحصائية المختلفة لتحليل البيانات المتوفرة في العينة.

5- استخلاص النتائج ووضع التوصيات:

الاستناد على التحليل الإحصائي لبيانات العينة لوصف النتائج حول المجتمع ومن ثم اقتراح حلول للمشاكل ووضع التوصيات المختلفة. ومن كل ما ورد يتضح أن علم الإحصاء يعتمد على المفاهيم والأسس العلمية التالية:

المجتمع population:

والذي يسمى عادة بالمجتمع الإحصائي statistical population والذي يتألف من مجموعة من الوحدات elements أو الأفراد observations التي تخص دراسة معينة.

أو هي تلك المفردات تحت الدراسة أو البحث ويهدف تعريف المجتمع الإحصائي إلى تعيين الحدود الصريحة لعملية جمع البيانات و لعملية الاستقراء أو الاستنتاجات التي يمكن الحصول عليها من الدراسة.

ويمكن ان تكون عناصر المجتمع أفراد أو عائلات أو موظفين أو مجموعة من الطلاب... الخ. يجب ان تكون معرفة محددة بحدود الزمن بحيث يستطيع الباحث معرفة انتماء أي عنصر من عدم انتمائه لذلك المجتمع وهذه المفردات المتكون منها المجتمع تسمى بالوحدات الإحصائية statistical element ولذلك فإن المجتمع الإحصائي أو المجتمع يتمثل بعدد الوحدات الداخلة في المجتمع والذي يدعى بحجم المجتمع ويرمز له بالرمز N .

العينة Sample:

وهي مجموعة جزئية من وحدات المجتمع. وأن تمثل حجم العينة بحيث $n < N$ وأن المعلومات المتوفرة في العينة هي تلك المتوفرة في المجتمع. وبذلك نضمن من خلال الإحصاء الاستدلالي ان النتائج التي يتم التوصل إليها من تحليل المعلومات في العينة ستعكس بصورة صحيحة تلك التي يمكن الحصول عليها من تحليل المعلومات من المجتمع.

على سبيل المثال عند القيام بدراسة حالة معينة عن طلبة جامعة عمان الأهلية فإننا نأخذ مجموعة صغيرة من الطلبة بدلا من القيام بالدراسة على جميع الطلبة، فهذه المجموعة الصغيرة تسمى عينة Sample ويمكن أن تكون فقط 500 وحدة " طالبا " من مجموع 5000 " طالباً " .

القياسات Measurements:

وحدات القياس للبيانات Measurement Scale:

التحليلات والحسابات الإحصائية تحتاج إلى تحديد نوع المقياس بالنسبة للبيانات حسب التحليل الإحصائي.

وهذه البيانات أو قيم المتغيرات هي عبارة عن قياسات Measurement تقيس تلك الظواهر أو تمثل قيم تلك المتغيرات. ومن وحدات القياس المستخدمة:

1- وحدات القياس الاسمية أو الفئات The Nominal or Categorical Scale:

ومن الاسم أنها تخص تسمية Naming المفردات أو تصنيفها Classifying حسب عدد من التصنيفات المختلفة مثل التصنيف حسب الجنس - ذكر أو أنثى - أو التصنيف حسب الحالة الاجتماعية - متزوج أو غير متزوج - أو التصنيف حسب الحالة الصحية - مريض أو غير ذلك - وفي هذه الحالة الترتيب غير مهم. وتعتبر من انطباق وحدات قياس البيانات المزدوجة Binary Data لتكون ما يسمى بجداول التوافق Contingency Tables والتي تستخدم لعرض البيانات الخاصة بنوعين من التصنيفات.

2- وحدات القياس المرتبة The Ordinal Scale:

عندما لا تصنف البيانات وفقا لتصنيفات مختلفة وحسب بل هناك ترتيب في قياسها حسب مؤشرات معينة عندئذ نقول أن القياس قد تم تحت وحدات القياس المرتبة Ordinal Scale فمثلا حالة المريض يمكن أن تصنف بتحسّن قليل أو تحسّن جيد أو تحسّن أفضل، وفي هذه الحالة يتم القياس حسب تصنيف المرتبة. وكذلك ذكاء الأطفال يمكن أن يصنف يفوق المتوسط أو أقل من المتوسط وهكذا، أي أن الترتيب مهم ولكن الفروقات بين الرتب غير متساوية. وفي جميع الأمثلة المماثلة فإن المفردات التي تقع ضمن تصنيف معين تعتبر متساوية ولكنها قد تكون أقل أو أفضل من المفردات التي تقع ضمن تصنيف آخر، وبذلك نعني أن المفردات ترتب Rank or Order من الأقل إلى الأكثر فمثلا ترتيب الناجحون حسب الرتب، الأول، الثاني، الثالث... أو المتسابقون في مسابقة معينة أيضا، الأول، الثاني،... الخ.

ولكن الفرق في الحالة الأولى يمكن أن يكون الفرق بين الأول والثاني يختلف بشكل كبير عن الفرق بين الثاني والثالث فهنا الفروق غير متساوية وليس ضروريا. وكذلك في الحالة الثانية أيضا الفرق غير ضروري.

3- وحدات القياس بالفترات The Interval Scale:

تحت هذا النوع من وحدات القياس نستطيع إضافة لترتيب القياسات تحديد المسافات بين أي نوعين من القياسات. فمثلا نقول بأن الفرق بين درجة حرارة 80°F و 90°F هو نفسه الفرق بين درجة حرارة 30°F و 40°F ، وعند قياس درجة الحرارة فإن

وحدة القياس هنا هي الدرجة ونقطة مقارنة تختار عشوائيا لتكون مثلاً نقطة الصفر والتي لا تعني انه ليس هناك درجة للحرارة ولهذا فهذا القياس خالي من الصفر. والقياس هنا يعتبر قياساً كمياً Quantitative Scale.

4- وحدات القياس بالنسبة The Ratio Scale:

وحدات القياس هنا تتميز بحقيقة تساوي النسب وكذلك تساوي الفترات، وتشابه وحدات القياس بالفترات السابق ذكرها من حيث مبدأ النقطة الصفرية، غير أنها تؤكد حقيقة أن النقطة الصفرية هنا تعني عدم وجود قيمة معينة للمفردة، فمثلاً الأوزان حيث لا نقول فقط بأن الفرق بين 100kg و 50kg فحسب بل أن الشخص الذي وزنه 100 kg يقال عنه اثنان مرتين من الشخص الذي وزنه 50kg. وغالباً ما لا يكون هناك تمييز بين وحدات القياس بالفترات والنسب.

ويلاحظ أن المتغيرات النوعية عادة ما تعرض بأحد الشكلين المصنفة أو المرتبة، أما المتغيرات الكمية فعادة ما تعرض بأحد الشكلين الفترات أو النسبة.

1-2 طبيعة البيانات:

الصفة التي تتغير من شخص إلى آخر أو من مفردة إلى أخرى تسمى ظاهرة Observation أو المتغير Variable ويرمز له (X) ولكل مفردة أو مشاهدة منها ترمز بالرمز $i = 1, 2, \dots, n$ (xi) فمثلاً عند دراسة أطوال مجموعة من الطلاب في جامعة عمان الأهلية فإننا نرمز لصفة الطول بالرمز X وطول أي طالب أو طالبة بالرمز (xi) وهذه تسمى بالمشاهدة أو المفردة observation ومن هنا يمكن تقسيم المتغيرات Variables إلى نوعين:

1- المتغيرات الوصفية أو النوعية Qualitative Variables:

وهي تلك الظواهر أو الصفات التي لا يمكن قياسها مباشرة بالأرقام العددية مثل صفة لون العين (أزرق، أخضر، أسود...) الحالة الاجتماعية (غني، متوسط، فقير، معدوم) وكذلك صفة الجنس، إلى غير ذلك من الصفات.

2- المتغيرات الكمية Quantitative variables:

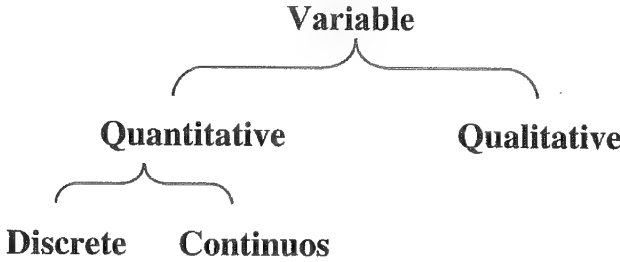
وهي تلك الظواهر او الصفات التي يمكن قياسها مباشرة بأرقام عددية مثل صفة الطول والوزن وكمية المحصول... الخ. ويمكن تقسيم المتغيرات الكمية إلى قسمين:

أ- المتغيرات المنفصلة Discrete Variables:

المتغير المنفصل هو المتغير الذي تأخذ المشاهدة او المفردة فيها قيما متميزة عن بعضها ومن أمثلتها رمي حجر النرد او رمي قطعة النقد وكذلك من الأمثلة الأخرى عدد أفراد الأسرة أو عدد الوحدات الإنتاجية في مصنع ما، بصورة عامة فإن كل البيانات التي نحصل عليها من العد Counting تعتبر بيانات لمتغير منفصل.

ب- المتغيرات المستمرة Continuous Variables:

وهي تلك المتغيرات التي تأخذ مدى معين او مجال معين من القيم ومن أمثلتها طول الطلبة في جامعة ما تتراوح ما بين 140 سم و 175 سم. يتبين من أعلاه أن العلاقة بين المتغيرات تكون كالآتي:



وبذلك فإن المعلومات التي تجمع وترتب وتحلل من قبل الإحصائي تدعى بالبيانات Data لذا يمكن تعريفها: بأنها المعلومات التي نحصل عليها من قيم المتغير وان البيانات الكمية هي البيانات التي نحصل عليها من قيم المتغير الكمي أما البيانات النوعية فنحصل عليها من قيم المتغير النوعي. أما البيانات المتقطعة نحصل عليها من قيم المتغير المتقطع، أما البيانات المستمرة نحصل عليها من قيم المتغير المستمر. وكأمثلة على ذلك فإن صنف الدم لشخص هي بيانات نوعية حصلنا عليها م قيمة المتغير النوعي صنف الدم. أما عدد الأسر في عمان فهو بيانات كمية متقطعة لأننا حصلنا عليها من قيمة المتغير الكمي المتقطع عدد الأسر وهكذا.

1-3 جمع البيانات Data collection:

يتم جمع البيانات عادة باستخدام الاستمارة الإحصائية، ومن طرق جمع البيانات:

1- المقابلة الشخصية:

يقوم الباحث بتوجيه الأسئلة للأشخاص أو المفردات وتدوين الإجابات في الأماكن المخصصة ولهذا الأسلوب مميزات حيث يمكن جمع أكبر قدر ممكن من الإجابة الصحيحة وكذلك تكون نسبة الاستجابة للمعلومات مرتفعة.

2- المراسلة:

يتم باستخدام الاتصال الهاتفي أو البريدي بحيث يذكر الباحث بكتاب مرفق أغراض البحث ومدى أهمية التعاون لإنجاح البحث. أما الاتصال الهاتفي يستخدم بدلا من المواجهة الشخصية لتقليل التكاليف وتوفير الوقت.

3- التسجيل المباشر:

وهنا يقوم الباحث بالانتقال إلى موقع العمل ومشاهدة الأشياء أو الأشخاص مثلا البحوث التي تجري على السير أو المرور في منطقة معينة، ظاهرة التهجين لبعض السلالات النباتية أو الحيوانية.

الاستمارة الإحصائية / استمارة استبيان:

تتألف عادة من صفحة أو عدة صفحات تحتوي على الأسئلة التي يريد الباحث إجابات عنها. يترك فراغ عند كل سؤال لتسجيل الإجابة عليها. وهنا يختلف تصميم الاستمارة من باحث إلى آخر وحسب موضوع البحث، إلا أن هناك شروطا عامة يجب مراعاتها عند تصميم الاستمارة منها :

- 1- أن تكون الأسئلة مختصرة وواضحة ولا تحتمل أكثر من تفسير.
- 2- أن يكون عددها أقل ما يمكن.
- 3- تفضل الأسئلة ذات الإجابة بنعم أو لا أو ذات الإجابات المتعددة واختيار إجابة منها أو ذات الإجابة المختصرة.

- 4- ان تكون الإجابات قابلة للتصنيف و التبويب. ومن أسس التصنيف:
- أساس نوعي: (ذكور وإناث).
 - أساس جغرافي: المناطق او المحافظات او حسب (ريف و حضر).
 - أساس زمني: تصنيف المهاجرين حسب سنة دخولهم.
 - أساس مشترك: حسب معيارين مثلاً (ريف-ذكور)، (ريف-إناث)، (حضر-ذكور) (حضر-إناث).

ويجب الإشارة إلى انه من الضروري تجربة الاستمارة بعد تصميمها وذلك بهدف اكتشاف النواقص ونقاط الضعف فيها وتعديلها إذا لزم الأمر. كما انه ومن المهم الإشارة وبوضوح في بداية الاستمارة إلى سرية المعلومات وإلى نوع وأهمية الدراسة التي يقوم بها، مما يشجع الوحدة الإحصائية الدقيقة في الإجابة.

تطبيقات SPSS

لتشغيل البرنامج الإحصائي SPSS نختار start ثم Programs ثم نختار SPSS وستفتح الشاشة على لوحة البيانات، وتكون اللوحة مقسمة أفقياً وعمودياً إلى مجموعة من المستطيلات (الخلايا Cells) تمثل الأعمدة فيها المتغيرات وتمثل الأسطر فيها المفردات المختلفة ضمن كل متغير. وبذلك يمكن ادخال قيم المتغيرات الواحد بعد الآخر.

1-4 عرض البيانات الإحصائية:

البيانات الأولية Row Data الخاصة بالدراسة. لا يمكن تفسيرها بشكل ملائم و الاستفادة منها وهي بهذه الصورة لذلك يلجأ الباحث إلى وضع تلك البيانات بشكل جداول مبسطة او التعبير عنها بالرسوم البيانية لكي يسهل دراستها وتحليلها. ومن طرق عرض البيانات نذكر:

1-4-1 العرض الجدولي والتوزيع التكراري Tabular Presentation and Frequency Distribution:

وهي عبارة عن عرض البيانات التي تم جمعها و وصفها في جداول منتظمة لذا فهناك نوعان من الجداول وهما الجداول البسيطة والجداول أو التوزيعات التكرارية والتي سيتم شرحها في الصفحات القادمة كما يلي:

1- الجداول البسيطة Simple Tables:

وهي تلك الجداول التي تتوزع فيها البيانات حسب صفة واحدة ويتألف عادة من عمودين، الأول يمثل الظاهرة Observation، والثاني يمثل عدد المفردات التابعة لكل مشاهدة وما يسمى بالتركرارات Frequency، ومن هذا المنطلق يمكن تسمية الجداول بجداول التوزيع التكراري Frequency Distribution Table.

أما الطريقة الأساسية لبناء التوزيع التكراري فهي عبارة عن تقسيم مدى قيم البيانات والتي تمثل المجموعة الكبيرة إلى فئات تمثل مجموعات صغيرة مختلفة ثم يتم حصر عدد البيانات الواقعة ضمن كل فئة. ويمكن استخدام جداول التوزيع التكراري في حالة البيانات الوصفية وكذلك في حالة البيانات الكمية المتقطعة والمستمرة Discrete and Continuous.

والمثال التالي يبين وصف البيانات المتقطعة وكيفية جدولتها.

مثال (1)

كُون جدول تكراري للبيانات التالية والتي تمثل الحالة الاجتماعية لـ 25 فردا.

أعزب	أرمل	مطلق	متزوج	مطلق	متزوج
مطلق	أرمل	أعزب	متزوج	أعزب	أعزب
متزوج	مطلق	أرمل	أعزب	متزوج	أعزب
أرمل	أعزب	متزوج	مطلق	أعزب	أعزب
أعزب					

الحل

نكون أولا جدولا مكون من ثلاث أعمدة العمود الأول يمثل الظاهرة Observation والعمود الثاني لتفريغ البيانات أما العمود الثالث فيمثل التكرار لكل مشاهدة او صفة وما يسمى Frequency وبذلك فإن جدول التوزيع التكراري للحالة الاجتماعية سيكون كالآتي:

التكرارات	توزيع البيانات	الحالة الاجتماعية
6	IIII I	متزوج
5	IIII	مطلق
10	IIII IIII	أعزب
4	IIII	أرمل
25		المجموع

2- الجداول أو التوزيعات التكرارية Frequency Distributions:

عندما يكون عدد المفردات قليلا او صغيرا يمكن عندئذ ملاحظة كل منها على انفراد أما عندما يكون عدد المفردات كبيرا فمن الأسهل و الأفضل استخدام ما يسمى بالجدول التكراري Frequency Table. ويتم ذلك بتجميع او توزيع البيانات كمجموعة كبيرة من القيم إلى اكثر من مجموعة او ما يسمى بالفئة Class او المجال Interval.

ولبناء جدول التوزيع التكراري علينا ملاحظة ما يلي:

أ- تحديد المدى Range:

والذي يمثل الفرق بين اكبر واصغر قيمة في المجموعة.

ب- اختيار عدد الفئات Number of Classes:

يمكن القول انه ليست هناك قاعدة عامة ولكن يمكن اختيار ذلك العدد من الفئات و الذي يتناسب مع حجم البيانات والأهداف المتوخاة من التحليل و يعتمد أيضا على خبرة الباحث. ويمكن أن نختار عدد الفئات فرضيا على أن لا تقل عن 5 وألا تزيد عن 15 وذلك تبعا لطبيعة البيانات وعدد مفرداتها.

ج- تحديد ما يسمى بعرض الفئة او مداها Class Width:

ويعتمد ذلك على خبرة الباحث و هيئة البيانات، وبشكل عام نستخدم عرض يسهل الحسابات المعتمدة عليها كأن نستخدم عرض فئة 5 او 10 وحدة وهكذا. ويمكن الاستناد على العلاقة التالية لتحديد ذلك العدد كآلاتي:

$$Classeswidth(w) = \frac{Range}{Numberofclasses} \text{ (مقربة إلى أقرب عدد صحيح أكبر)}$$

د- حدود الفئات Class Limits:

تبدأ حدود الفئات بأصغر قيمة او اصغر من ذلك بقليل و التي تمثل الحد الأدنى للفئة الأولى و تنتهي الفئة الأخيرة بالحد الأعلى و الذي يمثل اكبر قيمة او اكبر من ذلك بقليل وبذلك فإن كل فئة لها حدين ها الحد الأدنى Lower Cut Point و يرمز له L و الحد الأعلى Upper Cut point و يرمز له U وتكتب بالشكل العام التالي:

$$[L, U) \text{ أو } (الحد الأعلى, الحد الأدنى]$$

وتعتبر هذه الصيغة من انسب الصيغ المستخدمة لكتابة الفئات و تعني ان الحد الأدنى يدخل ضمن الفئة المعنية بينما لا يدخل الحد الأعلى فيها وإنما سيدخل في الفئة التي تليها. وبذلك يلاحظ ان عرض الفئة يمكن حسابه باستخدام:

$$\text{Class width (w)} = U - L$$

هـ- مراكز الفئات Class Marks او Class Midpoint:

فإنها تمثل المتوسط للحد الأعلى والحد الأدنى لكل فئة أي ان مركز الفئة والذي يرمز له بالرمز xi يمكن حسابه باستخدام العلاقة:

$$Xi = \frac{Li + Ui}{2}$$

أما عن الحدود الدنيا والحدود العليا الفعلية للفئات class boarders فيتم ذلك بطرح نصف من قيمة الحد الأدنى لكل فئة وإضافة نصف للحد الأعلى لكل فئة.

و- تفريغ البيانات وإيجاد عدد التكرارات لكل فئة class frequency:

تشكيل الجدول التكراري يعني توزيع المشاهدات او البيانات الموجودة على العدد المحدد المناسب من الفئات ومن ثم تحديد عدد المشاهدات او البيانات المفرغة في كل فئة.

يتم تسجيل القيم الواحدة بعد الأخرى في الفئة الخاصة بها على شكل إشارة او علامات أولاً ثم ترجمتها إلى أعداد (كما سيتم توضيح ذلك لاحقاً) و بعدها يتم جمع هذه التكرارات للتأكد من المجموع و الذي يمثل حجم العينة. العدد المعين من المشاهدات او البيانات والذي يخص فئة معينة يسمى بتكرار الفئة class frequency و يرمز له fi والذي يمثل تكرار الفئة i.

وبذلك فإن الشكل العام للجدول التكراري يتكون من 3 أعمدة رئيسية وهي عمود للفئات وعمود للقيم الموزعة (التكرار بالإشارات أو العلامات) وعمود يمثل مجموع عدد القيم (التكرار). ويضاف عادة عمود أول يشير إلى تسلسل الفئات. سوف نتطرق إلى التوزيع التكراري في حالة البيانات المتقطعة و البيانات المستمرة بأخذ مثال لتوضيح كل حالة.

مثال (2)

فيما يلي بيانات عن عدد أفراد الأسرة في عينة حجمها 25 أسرة. كَوّن جدول تكراري لهذه البيانات.

2	4	5	7	7	6	4	2	3	3
5	4	4	3	5	6	3	4	4	5
7	4	5	3	6					

الحل

نلاحظ هنا بان البيانات متميزة عن بعضها بالشكل 2،...، 3 ولذلك نرتب البيانات تصاعدياً أولاً لتمثل العمود الأول أما العمود الثاني فيبين تفريغ البيانات بالعلامة أما العمود الثالث فيمثل تكرارات كل أسرة.

جدول (1)

التوزيع التكراري لـ 25 أسرة حسب الحجم

حجم الأسرة	تفريغ البيانات	التكرارات (عدد الأسر)
2	II	2
3	IIII	5
4	II IIII	7
5	IIII	5
6	III	3
7	III	3
المجموع		25

مثال (3)

البيانات التالية تمثل درجات 50 طالبا من طلبة جامعة عمان الأهلية في مادة الإحصاء، كوّن جدول توزيع تكراري.

76	75	95	96	92	84	80	64	65	59	52	50	44
40	70	68	65	66	76	77	98	84	83	57	56	49
47	45	58	58	86	85	80	90	90	87	86	74	69
72	73	88	66	67	76	58	89	64	70	54		

الحل

لعمل الجدول التكراري نتبع الخطوات التالية:

أ- المدى $\text{range} = 98 - 40 = 58$

ب- عدد الفئات كما ذكرنا سابقا ونفرض عدد الفئات تساوي 6.

ج- إيجاد طول الفئة $w = \frac{\text{Range}}{\text{Number of classes}}$

$$w = \frac{58}{6} \approx 10$$

وللحفاظ على توازن الفئات قربنا طول الفئة إلى 10.

د- حدود الفئة: نعين الحد الأدنى لأول فئة ويجب أن يكون هذا الحد مساويا لأصغر قيمة في البيانات أو اقل منها بقليل فالحد الأدنى من هذا المثال هو 40، وبعد تعيين الحد الأدنى للفئة نعين الحد الأدنى الفعلي لتلك الفئة وهو عبارة عن الحد الأدنى ناقصا نصف وحدة من الوحدات التي ضربت إليها الأعداد في البيانات فإن الحد الأدنى الفعلي للفئة الأولى هو 39.5. وبعد ذلك نحدد الحد الأعلى للفئة الأولى بإضافة طول الفئة أي (40 + 10) و لكن لو نظرنا إليها من جانب الحد الأعلى الفعلي لتلك الفئة فيجب طرح

نصف الوحدة فيكون 495 و يمكن كتابة حدود الفئة لسهولة العمل 40 و اقل من 50 أو (-40).

هـ- تعيين الحدود الدنيا و العليا لجميع الفئات الباقية و ذلك بإضافة طول الفئة لكل حد.

و- تفرغ البيانات في الجدول ثم نجد التكرارات المقابلة لكل فئة.

ز- إيجاد مركز الفئات xi.

وأخيرا فإن جدول التوزيع التكراري لعلامات الطلبة سيكون:

جدول (2)

التوزيع التكراري لدرجات طلبة جامعة عمان

عدد الطلبة التكرار (f)	التكرار بالإشارات	مراكز الفئات (xi)	الحدود الضمنية للفئات	فئات الدرجات
5	IIII	45	39.5-49.5	40-
9	IIII III	55	49.5-59.5	50-
9	IIII III	65	59.5-69.5	60-
10	IIII IIII	75	69.5-79.5	70-
11	IIII IIII I	85	79.5-89.5	80-
6	IIII I	95	89.5-99.5	90-100
50				المجموع

3- الجداول الشائبة Contingency tables:

وهي تلك الجداول التي تتوزع فيها البيانات حسب صفتين او ظاهرتين في نفس الوقت. وبذلك فإن الجدول الشائبي يتكون من الصفوف والتي تمثل فئات أو مجاميع إحدى الظاهرتين والأعمدة التي تمثل فئات أو مجاميع الصفة الأخرى إلى المربعات الناتجة من الصفوف والأعمدة فتحوي على عدد المفردات او التكرارات المشتركة بين الظاهرتين. والمثال التالي يبين وصف الجداول الشائبة.

مثال (4)

البيانات التالية تمثل الحالة الاجتماعية حسب الجنس إلى 50 شخصاً.

جدول رقم (3)
الحالة الاجتماعية حسب الجنس.

الحالة الاجتماعية الجنس	أعزب	متزوج	مطلق	أرمل	مجموع
ذكور	5	12	7	4	28
إناث	7	8	1	6	22
	12	20	8	10	50

مثال (5)

جدول التوزيع التكراري الشائي التالي يمثل عدد من طلبة جامعة عمان الأهلية مصنّفين حسب صفتي الطول والوزن.

جدول رقم (4)
الطلاب حسب الوزن والطول

الوزن الطول	50	60-	70-80	المجموع
120-	12	4	0	16
140-	3	25	6	34
160-180	5	6	14	25
المجموع	20	35	20	75

1-4-2 التوزيع التكراري النسبي Relative Frequency distribution:

وهو جدول يبين الأهمية النسبية لكل فئة، وإن التكرار النسبي ويرمز له Rel-fi و يمثل قسمة تكرار كل فئة إلى مجموع التكرارات:

$$Rel.fi = \frac{fi}{\sum fi = n}$$

وبذلك فإن الجدول الذي يعرض الفئة أو مراكز الفئة مع التكرار النسبي يسمى جدول تكراري نسبي. وعليه يمكن حساب التكرار النسبي للمثال رقم (1). وإذا ضربنا التكرار النسبي $\times 100\%$ نحصل على التكرار النسبي المئوي. ويجب ملاحظة أن مجموع التكرارات النسبية دائماً يساوي واحد.

مثال (6)

احسب التكرار النسبي للمثال السابق (الأجور)

جدول رقم (5)

جدول التكرار النسبي لدرجات 50 طالباً

فئات الدرجات	عدد الطلبة (fi)	Rel-fi
40-	5	5/50
50-	9	9/50
60-	9	9/50
70-	10	10/50
80-	11	11/50
90-100	6	6/50
المجموع	50	1

1-4-3 التوزيع التكراري المتجمع Cumulative frequency distribution:

سبق وتحدثنا عن جداول التوزيع التكراري للفئات الذي يبين توزيع قيم الظاهرة. ولكن في بعض الأحيان قد نحتاج إلى معرفة عدد القيم او المفردات (المشاهدات) التي تقل أو تزيد عن قيمة معينة. لذا سوف نتطرق إلى جانب التوزيعات التكرارية المتجمعة للفئات وهناك نوعان من هذه التوزيعات المتجمعة وهما:

1- التوزيع التكراري المتجمع الصاعد $CF \uparrow$ Increasing Cumulative Frequency:

ويرمز له عادة إما ucf أو $CF \uparrow$ وجدول التوزيع التكراري المتجمع الصاعد يتكون عادة من عمودين، العمود الأول يمثل الحدود والعمود الثاني التكرار المتجمع الصاعد الذي يكون حسابه كما يلي:

(أ) التكرار المتجمع الصاعد للفئة الأولى يمثل نفس تكرار الفئة الأولى لأن تكرار الفئة السابقة للفئة الأولى = صفر. أي أن:

$$F_1 \uparrow = f_1$$

(ب) التكرار المتجمع الصاعد للفئة الثانية يساوي التكرار المتجمع الصاعد للفئة الأولى + تكرار الفئة الثانية أي:

$$F_2 \uparrow = F_1 \uparrow + f_2$$

(ج) هكذا نستمر بإضافة تكرار الفئة التالية للتكرار المتجمع الصاعد إلى أن نصل إلى آخر فئة حيث يمثل مجموع التكرارات الكلية أي أن:

$$F_n \uparrow = \sum_{i=1}^n f_i$$

كذلك يمكن استخراج التكرار المتجمع الصاعد النسبي لكل فئة وذلك بقسمة التكرار التجميعي الصاعد لكل فئة على مجموع التكرارات بحيث يكون التكرار النسبي الصاعد للفئة الأخيرة يساوي واحد.

والجدول التالي يبين التكرار المتجمع الصاعد و التكرار المتجمع النسبي باستخدام نفس المثال السابق.

جدول (6)

جدول التكرار المتجمع الصاعد والتكرار المتجمع الصاعد النسبي للدرجات

فئات الدرجات	عدد الطلبة (f _i)	الحدود الفعلية للفئات	CF ₁ ↑	Rel-CF ₁ ↑
40-	5	less than 50	5	5/50
50-	9	less than 60	14	14/50
60-	9	less than 70	23	23/50
70-	10	less than 80	33	33/50
80-	11	less than 90	44	44/50
90-100	6	less than 100	50	50/50
المجموع	50			

2- التوزيع التكراري المتجمع النازل CF₁ ↓ : Decreasing Cumulative Frequency

ويرمز له أيضا LCF أو CF₁ ↓ وهو التوزيع الذي يعطي عدد المفردات التي تزيد قيمها عن الحد الأدنى لفئة معينة. ويتكون هذا التوزيع من عمودين، عمود يمثل الفئات والثاني يمثل التكرار المتجمع النازل. والذي يمكن حسابه كما يلي:

(أ) التكرار المتجمع النازل للفئة الأولى يتمثل بمجموع التكرارات الكلية أي أن :

$$F_1 \downarrow = \sum_{i=1}^n f_i$$

(ب) التكرار المتجمع النازل للفئة الثانية يساوي التكرار المتجمع النازل للفئة الأولى مطروحا منه التكرار المناظر له أي تكرار الفئة الأولى أي أن:

$$F_2 \downarrow = \sum_{i=1}^n f_i - f_1$$

(ج) التكرار المتجمع النازل للفئة الثالثة F₃ ↓ سوف يكون :

$$F_3 \downarrow = F_2 \downarrow - f_2 = \sum_{i=1}^n f_i - f_1 - f_2$$

(د) وهكذا نستمر بالتنازل إلى آخر فئة للحصول على التكرار المتجمع النازل له F_n يساوي تكرار الفئة الأخيرة f_n . وكذلك الحال بالنسبة إلى التكرار المتجمع النازل النسبي يشبه التكرار النسبي الصاعد في عملية استخراجها. والمثال التالي يبين التكرار المتجمع النازل و التكرار التجميعي النسبي وكذلك التكرار التجميعي الصاعد.

جدول (7)

جدول التكرار المتجمع النازل والتكرار المتجمع النازل النسبي للدرجات

فئات الدرجات	عدد الطلبة (f_i)	الحدود الفعلية للفئات	$CF_1 \downarrow$	$Rel-CF_1 \downarrow$
40-	5	40& more	50	50/50
50-	9	50& more	45	45/50
60-	9	60& more	36	36/50
70-	10	70& more	27	27/50
80-	11	80& more	17	17/50
90-100	6	90& more	6	6/50
المجموع	50			

مثال (7)

البيانات التالية تمثل الأجر الأسبوعي ل 40 عاملا بالدينار الأردني.

المطلوب:

- 1- أوجد التوزيع التكراري المتجمع الصاعد.
- 2- أوجد التوزيع التكراري المتجمع النازل.
- 3- عدد العمال الذين يتقاضون أجرا 60 دينار فأكثر.
- 4- عدد العمال الذين يتقاضون أجرا اقل من 50 دينار.

نحسب التكرار المتجمع الصاعد و النازل أولاً بشكل الجدول التالي:

جدول (8)

التكرار المتجمع الصاعد والنازل للأجر الأسبوعي

فئات الأجر	عدد العمال (F _i)	CF _i ↑	CF _i ↓
30-	3	3	40
40-	1	4	37
50-	8	12	36
60-	10	22	28
70-	7	29	18
80-	7	36	11
90-100	4	40	4
المجموع	40		

عدد العمال الذين يتقاضون أجراً 60 دينار فأكثر هم

$$10+7+7+4=28$$

ويمكن حسابه مباشرة من التكرار التجمعي النازل الذي يقابل فئة (60-).

أما عدد العمال الذين يتقاضون أجراً أقل من 50 دينار فهم:

$$3+1=4$$

ويمكن حسابه مباشرة من التكرار التجمعي الصاعد الذي يقابل الفئة (40-).

لتطبيقات SPSS:

ويمكن عمل هذا الجدول من خلال الإيعازات التالية للبرنامج SPSS وهو الخيار Analyze من القائمة الرئيسية ثم الخيار Descriptive وبعد هذا الخيار Frequencies ويحدد المتغير Variable المراد إيجاد الجدول التكراري له.

1-5 التمثيل البياني للتوزيعات التكرارية:

Graphical Presentation for Frequency Distributions

يمكن توضيح البيانات بشكل مناسب وذلك باستخدام الرسوم البيانية والصور والأشكال الهندسية بحيث تساعد القارئ على سهولة فهم المعلومات الواردة واستيعاب قيم الظاهرة وبالتالي يمكن مقارنتها مع بعضها. ويمكن استخدام مختلف الرسومات البيانية في علم الإحصاء بحيث يمكن رسم البيانات المعروضة بشكل توزيع

تكراري أما لعملية الرسم نستخدم المحاور Coordinates المحور السيني X- axis والذي يمثل الفئات. أما المحور الصادي Y- axis والذي يمثل التكرارات. وهناك أنواع مختلفة من الرسومات البيانية سوف نتطرق إلى بعض منها أولاً في حالة إذا كانت البيانات (وصفية) Qualitative Data وهي:

1- الأشرطة البيانية Bar Chart:

وهنا الرسم عبارة عن مستطيلات رأسية تتخذ قواعدها على المحور الأفقي لتمثل الظاهرة بينما ارتفاعها يمثل التكرارات لكل مفردة داخل الظاهرة بينما الارتفاع يمثل تكرار كل مفردة داخل الظاهرة و لرسم الأشرطة البيانية نحتاج إلى الخطوات التالية:

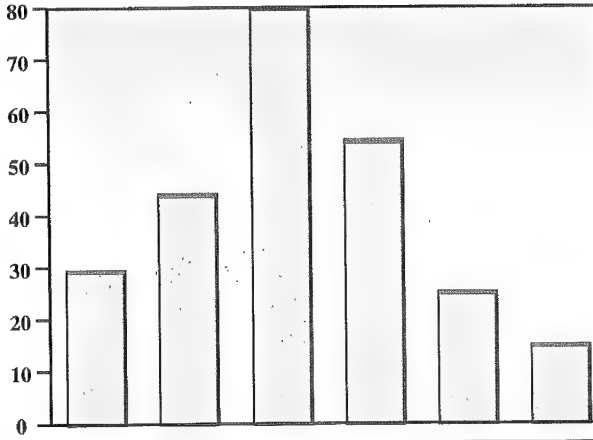
أ- ندرج المحور الأفقي إلى أقسام متساوية بمقياس رسم مناسب بحيث يشمل جميع البيانات فإن قاعدة المستطيل تمثل صفة الظاهرة الوصفية ويفضل ترك مسافة متساوية بين المستطيل والآخر.

ب- أما المحور العمودي فيمثل التكرارات أيضاً مقسمة بمقياس رسم مناسب بحيث كل صفة تقابل التكرار المناسب لها.

مثال (8)

الجدول التكراري التالي يمثل تخصصات مختلفة لعدد من الطلبة المسجلين فيها:

عدد الطلاب (f)	التخصصات (صفة)
30	علوم سياسية
45	علوم زراعية
80	علوم هندسية
55	علوم طبية
25	علوم اجتماعية
15	آداب
250	المجموع



الشكل (1) الأشرطة البيانية لعدد الطلاب حسب التخصص

لتطبيقات SPSS:

ويمكن عمل هذا الرسم البياني باستخدام البرنامج SPSS وكما يلي: أن نختار graphs من القائمة الرئيسية ومنها نحدد الخيار Bar ونختار الخيار simple ومن ثم نحدد المتغير variable المطلوب رسمه.

2- الرسم الدائري Pie Chart:

يتم عرض البيانات من خلال رسم دائري و المتمثل بمجموعة من الأجزاء داخل المجموع الكلي لهذه الأجزاء والذي يمثل زاوية الدائرة. 360 هذه الأجزاء تتمثل بقطاعات دائرية ويأخذ جزء من زاوية الدائرة لذا يجب اخذ الخطوات التالية بنظر الاعتبار:

أ- نحسب التكرار النسبي لكل صفة $Rel-f_i$

ب- نضرب التكرار النسبي $Rel-f_i$ بزاوية الدائرة للحصول على زاوية القطاع لكل صفة.

ج- يتم رسم الأجزاء ابتداء من أكبر مقطع من الزاوية بعد تقسيم الدائرة إلى أربع أقسام ليسهل عليه الرسم ويمكن إعطاء ألوان مختلفة لكل مقطع لتوضيح شكل او حجم الصفة التي تعود لها.

مثال (9)

بالرجوع لبيانات مثال (8) السابق ارسم Pie Chart للمعلومات

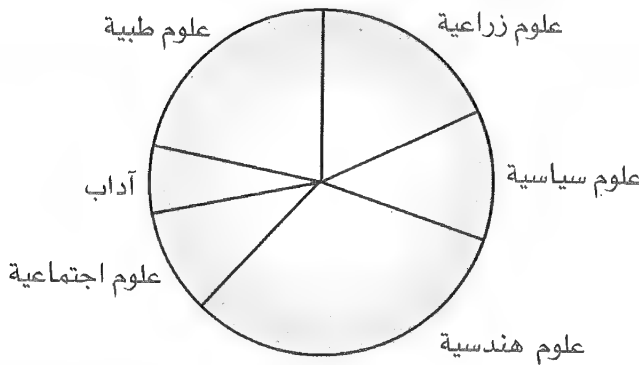
الحل

خطوات عمل الدائرة تعطينا الجدول التالي:

جدول رقم (9)
يبين دائرة بيانية لتوزيع الطلاب

الصفة	fi	Rel.fi	زاوية القطاع
علوم سياسية	30	0.12	$0.12 \times 360 = 43.2$
علوم زراعية	45	0.18	$0.18 \times 360 = 64.8$
علوم هندسية	80	0.32	$0.32 \times 360 = 115.2$
علوم طبية	55	0.22	$0.22 \times 360 = 79.2$
علوم اجتماعية	25	0.10	$0.10 \times 360 = 36.0$
آداب	15	0.06	$0.06 \times 360 = 21.6$
	250		

والمعلومات في الجدول تظهر الرسم البياني التالي:



الشكل (2) الرسم الدائري للتخصصات المسجلين عليها من قبل الطلاب

تطبيقات SPSS:

ويمكن عمل هذا الشكل البياني باستخدام البرنامج SPSS وكما يلي: نختار Graphs من القائمة الرئيسة ومنها نختار Pie وبعد ذلك نحدد نوع البيانات والمتغير المطلوب. أما في حالة البيانات الكمية فالتمثيل البياني لها يتم بإحدى الطرق التالية:

1- المدرج التكراري Frequency Histogram:

وهو نوع آخر من الرسوم الإحصائية مشابهة للأشرطة البيانية ويختلف عن الأشرطة البيانية أن المستطيلات متلاصقة والأساس يتخذ الحدود الفعلية للفئات بدلا من الاعتيادية. ويتكون المدرج من محورين متعامدين؛ حيث يمثل المحور الأفقي الفئات Class أما المحور العمودي الثاني فيمثل التكرارات Frequency. وهنا أيضا يجب أن تكون قواعد المستطيلات متساوية لأنها تمثل عرض الفئة Class Width. والمثال التالي يوضح الفكرة.

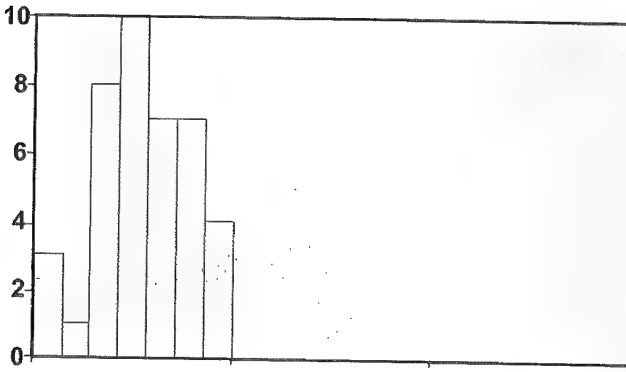
مثال (10)

الجدول التالي التكراري التالي يمثل الأجور الأسبوعية إلى 45 عاملا بالدينار الأردني:

جدول (10)
الأجر الأسبوعي لـ 40 عاملا

فئات الأجر	عدد العمال (f)
30-	3
40-	1
50-	8
60-	10
70-	7
80-	7
90-100	4
المجموع	40

المطلوب: ارسم مدرج تكراري.



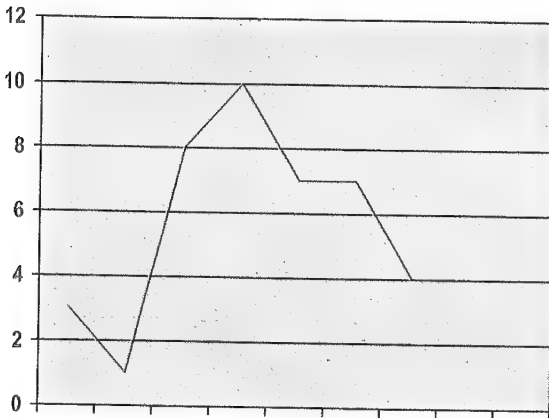
الشكل (3) مدرج تكراري للأجر الأسبوعي

تطبيقات SPSS:

ويمكن عمل هذا الشكل البياني باستخدام البرنامج SPSS وكما يلي: نختار Graphs من القائمة الرئيسية ومنها نختار Histogram ونحدد المتغير المطلوب.

2- المضلع التكراري frequency polygon:

طريقة أخرى لعرض البيانات الإحصائية ويمكن رسمها مباشرة من المدرج التكراري وذلك بتصنيف الأضلاع العلوية لمستطيلات المدرج التكراري ثم توصيل هذه النقاط ببعضها البعض بخطوط متكسرة كما هو موضح في الشكل أدناه وكذلك

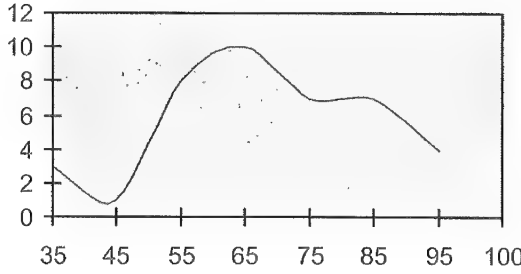


يمكن رسمها على المحورين مباشرة بأخذ مراكز الفئة على المحور الأفقي إلى المحور العمودي فيمثل التكرارات ونحدد جميع النقاط ونوصل فيما بينها بخطوط مستقيمة متكسرة. كما هو موضح في الشكل (4) لنفس المثال السابق:

الشكل (4) مضلع تكراري للأجر الأسبوعي

3- المنحنى التكراري frequency curve:

وهو عبارة عن منحنى يمر بمعظم النقاط التي يمثل إحداثها السيني مراكز الفئة وإحداثها الصادي تكرار تلك الفئة، وتكون مساحته مكافئة للمضلع التكراري و المساحة المحصورة تحت هذا المنحنى له الشكل (5) يوضح المنحنى التكراري للمثال السابق:



الشكل (5) المنحنى التكراري للأجر الأسبوعي

نظريات SPSS:

ويمكن عمل هذا الشكل البياني باستخدام البرنامج الإحصائي SPSS وكما يلي:

أن نختار Graphs من القائمة الرئيسة ومن هذا الخيار نختار Line ونحدد من هذا الخيار Simple ونختار ثالث خيار من البيانات وننقر على define لنختار المتغير variable ثم ننقر على ok.

وهنا أيضاً للتوزيع الطبيعي من خلال خيار الانحدار Regression ويمكن تحديد Line من خيار Plot.

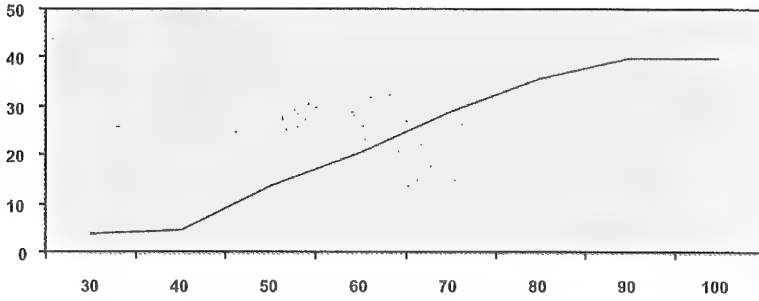
4- المنحنى المتجمع الصاعد والنازل Cumulative Frequency Curve:

وهو عبارة عن منحنى متصل بين نقاط واقعة فوق الحدود الحقيقة للفئات وعلى ارتفاع يمثل التكرار التجميعي سواء الصاعد او النازل حسب الخطوات التالية :
أ- ارسم محورين.

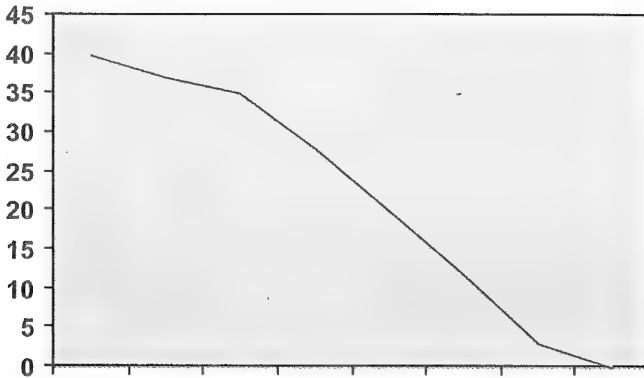
ب- ندرج المحور الأفقي أقسام متساوية تمثل الفئات كلها ويقسم العمودي إلى أقسام متساوية بحيث تمثل التكرار التجميعي $\sum f_i$.

ج- وضع نقاط أمام كل حد فئة ارتفاعها يعادل التكرار الصاعد او النازل لذلك الحد

د- نوصل تلك النقاط ببعضها البعض فيتكون لدينا منحنى متجمع صاعد وكذلك الحال في المنحنى المتجمع النازل كما في الشكلين التاليين.



الشكل (6) منحنى متجمع صاعد للأجر الأسبوعي



الشكل (7) منحنى متجمع نازل للأجر الأسبوعي

1-6 أشكال التوزيعات التكرارية Shapes of Distributions:

لاحظنا من المبحث السابق ومن رسم التوزيع أن هناك أشكال مختلفة تعطي فكرة قد تكون في كثير من الأحيان واضحة عن التوزيع ولذلك فإن البحث في أشكال التوزيعات التكرارية أصبح ضرورياً ومهماً لوصف البيانات. وبصورة عامة هناك أعداد لا نهائية من أشكال التوزيعات، ولكن هناك مقاييس أو معاملات عديدة يمكن

استخدامها لمعرفة التغير في الشكل وبدقة وسنبحث بعضاً منها وبالتفصيل لاحقاً. أما الآن فنكتفي بوصف خواص التوزيعات وصفاً إنشائياً مع أن هذا الوصف أقل وضوحاً من المقاييس العددية حيث أن هناك ثلاث خواص يجب معرفتها عند وصف البيانات هي الشكل والنزعة المركزية والتشتت.

وسيتم في هذا المبحث دراسة أشكال التوزيعات التكرارية من خلال تمييزها حسب أحد التمييزات التالية وهي:

1- التمييز بين التوزيعات المتماثلة والتوزيعات الغير متماثلة:

وبملاحظة الشكل (8) التالي نستطيع تحديد الفرق:



(أ) توزيعات متماثلة:

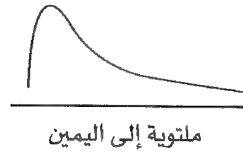
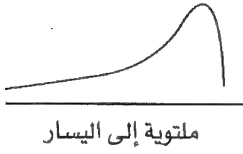


(ب) توزيعات غير متماثلة:

الشكل (8) أشكال التوزيعات المتماثلة والغير متماثلة

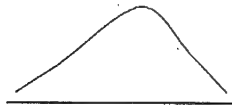
من الناحية النظرية فإن التوزيعات عادة ما تكون متماثلة ومنها التوزيع الطبيعي Normal distribution وهو الشكل الجرسى الظاهر في الفقرة (a) من الشكل (8) أعلاه والذي يعتبر أهم التوزيعات الإحصائية على الإطلاق (وستتم دراسته وبشكل وافي في الفصول القادمة)، أما في الحياة العملية فيوجد عدد قليل من التوزيعات المتماثلة ولكن يوجد كثير من التوزيعات التي تكون قريبة من التماثل.

أما التوزيعات التي يكون فيها عدم التماثل واضحاً فتسمى توزيعات ملتوية Skewed distributions وقد تكون ملتوية إلى اليمين Right Skewed أو ملتوية إلى اليسار Left skewed كما في الشكل (9) أدناه.



الشكل (9) التوزيعات الملتوية

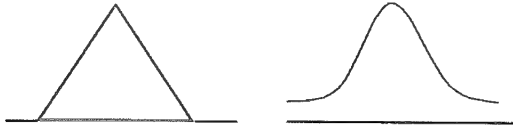
أما الشكل (10) التالي فيبين توزيعاً يعتبر معتدل الالتواء أو بسيط الالتواء أو قريب من التماثل:



الشكل (10) توزيع قريب من التماثل

2- التمييز بين التوزيعات بقمة واحدة أو متعددة القمم:

فيمكن ملاحظة الفرق بين نوعي التوزيعات بملاحظة الشكل (11) التالي:



(أ) توزيعات بقمة واحدة

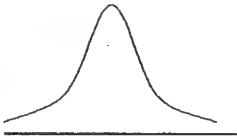


(ب) توزيعات بعدة قمم

الشكل (11) توزيعات بقمة أو أكثر

3- التمييز بين التوزيعات من حيث اتساعها (تضطحها):

ويمكن ملاحظة الفرق بين توزيعات باتساع صغير أو كبير من الشكل (12):



(ج) توزيع قليل التفرطح (مدبب)



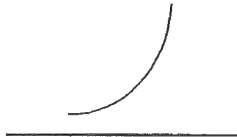
(ب) توزيع متوسط التفرطح



(أ) توزيع كبير التفرطح

الشكل (12) توزيعات بتفرطح مختلف

4- في كثير من الأحيان هناك تسميات معينة لبعض التوزيعات والتي تصف لنا التوزيع وصفاً دقيقاً. وأكثر ما تتضح هذه التسميات في الشكل (13):



(ب) توزيع بشكل J



(أ) توزيع مستطيلي (متجانس)

ويسمى أيضاً Uniform



(د) توزيع مثلثي



(ج) توزيع بشكل J المعكوس



(هـ) توزيع شكل الجرس

ويسمى التوزيع الطبيعي

الشكل (13) توزيعات بأسماء محددة

أسئلة

الفصل الأول

- 1- ما هو المقصود بالإحصاء الاستدلالي؟
- 2- إذا كانت مراكز الفئة لأعمار عدد معين من الطلبة هي:
18 21 24 27 30 فما هو:
 - 1- طول الفئة. 2- حدود الفئة لهذا التوزيع. 3- عدد الفئات.
- 3- صنف المتغيرات التالية إلى متغيرات وصفية أو كمية مبيناً فيما إذا كانت منقطعة أو مستمرة:
 - 1- عدد التلفونات في كل منزل.
 - 2- نوع التلفون.
 - 3- عدد المكالمات الخارجية في كل شهر.
 - 4- أطول مكاملة خارجية في الشهر بالدقائق.
 - 5- قيمة الفاتورة بالدينار الأردني للمكالمات الخارجية.
- 4- الجدول التالي يبين عدد المؤسسات الخدمية حسب النوع:

المؤسسات الخدمية	العدد
فنادق	3
موتيل	4
شقق فندقية	11
شقق للإيجار (مفروش)	12
أخرى	2

مثل البيانات بأفضل رسم ممكن.

5- البيانات التالية تمثل القيمة الدفترية لـ 50 سهماً في الأسواق المالية:

7	9	8	6	12	6	9	15	9	16
8	5	14	8	7	6	10	8	11	4
10	6	16	5	10	12	7	10	15	7
10	8	8	10	18	8	10	11	7	10
7	8	15	23	13	9	8	9	9	13

المطلوب:

- أحسب المدى.
 - عدد الفئات المناسبة.
 - كون جدول توزيع تكراري.
 - ارسم المدرج التكراري.
 - أوجد التكرار المتجمع الصاعد والنازل مع الرسم.
- 6- وزعت استمارة استبيان لـ 616 شخصاً وكانت الإجابات لاستطلاع معين كما يلي:

نعم	146
كلا	91
غير متأكد	58
غير مسؤول	123

المطلوب:

- رسم الأشرطة البيانية.
- الرسم الدائري.
- أيهما أفضل.

7- البيانات التالية تمثل الأجر الأسبوعي بالدينار الأردني لـ 30 عامل في مصنع الإسمنت:

47	48	40	43	45	39	45	50	46	52
42	47	52	35	38	32	52	33	50	43
55	51	49	50	37	31	39	43	60	60

كون جدول توزيع تكراري للأجور.

8- الجدول التالي يبين الحالة التعليمية لـ 200 شخص حسب الجنس:

الحالة التعليمية	لا يقرأ ولا يكتب	مؤهل متوسط	مؤهل إحصائي	مؤهل جامعي
ذكور	19	25	28	28
إناث	15	37	28	20

المطلوب:

- عرض هذه البيانات باستخدام الأعمدة البيانية.

- الرسم الدائري.

- إيجاد نسبة الذكور الذين لديهم مؤهل متوسط، ثم نسبتهم.

- إيجاد نسبة الإناث اللواتي لديهن مؤهل جامعي، ثم نسبتهم.

9- سحبت عينة عشوائية من إحدى شركات الغزل والنسيج، وكان إنتاجهم الأسبوعي بالوحدة موضح في الجدول التالي:

الجموع	55-60	50-	40	45-	35-	30-	25-	20-	كمية الإنتاج
250	20	27	36	44	55	32	22	14	عدد العاملين

المطلوب:

- رسم المدرج التكراري النسبي.

- إيجاد التكرار المتجمع الصاعد والنازل مع الرسم.

- إيجاد عدد العمال الذين ينتجون 30 وحدة وأكثر، ثم نسبتهم.

- إيجاد عدد العمال الذين ينتجون ما بين 35 و 55 وحدة ثم نسبتهم.

- إيجاد الحد الأعلى للإنتاج الذي يحققه 150 عاملاً.

10- البيانات التالية تمثل أعمار مجموعة من مرضى فقر الدم:

10	60	83	76	21	65	47
64	45	53	82	68	38	70
54	23	47	56	79	68	61
7	77	75	79	48	38	61
59	55	41	83	57	41	48

المطلوب:

- إيجاد عدد ثم نسبة المرضى اللذين أعمارهم 60 سنة فأكثر.

- إيجاد عدد ثم نسبة المرضى اللذين أعمارهم أقل من 30 سنة.

- كتابة البيانات بشكل جدول تكراري باستخدام أطول فئات 10 لكل منها ثم رسم الجدول بالطرق الممكنة.

- كتابة البيانات بشكل جدول تكراري متجمع صاعد وآخر متجمع نازل ورسمها.

11- البيانات الشائية التالية تمثل تصنيف مجموعة من طلبة الجامعة حسب العمر

والجنس كالآتي:

العمر	الجنس	العمر	الجنس	العمر	الجنس
22	M	21	F	29	F
23	M	28	F	20	M
19	F	21	F	18	F
21	M	24	F	21	M
21	F	24	F	26	M
21	F	21	F	24	F
20	F	23	M	19	M
20	F	20	F	25	M
20	F	19	M	19	F
22	F	24	F	23	M

المطلوب:

- كتابة البيانات بشكل جدول توافق.
- كتابة البيانات التي تخص العمر بشكل جدول تكراري باستخدام بداية الفترة الأولى 20 وعدد الفئات 4، ورسمه بالطرق الممكنة.
- كتابة البيانات التي تخص الجنس بشكل جدول ورسمه بالطرق الممكنة.
- كتابة البيانات بشكل جدول توافق.
- رسم بيانات العمر والجنس بشكل دائرة بيانية.
- رسم بيانات العمر بشكل دائرة بيانية.
- رسم بيانات الجنس بشكل دائرة بيانية.
- ايجاد عدد ثم نسبة الطلبة الذين تقل أعمارهم عن 20 سنة.
- ايجاد عدد ثم نسبة الطلبة الذكور الذين تقل أعمارهم عن 20 سنة.
- ايجاد عدد ثم نسبة الطلبة الإناث الذين تقل أعمارهم عن 20 سنة.

2

الفصل الثاني

مقاييس الترعة المركزية والنشت

Measures of Central Tendency and Dispersion

2-1 مقدمة

2-2 الوسط الحسابي

2-3 الوسيط

2-4 المنوال

2-5 الوسط الهندسي

2-6 الربيعيات والعشريات (المئينات) والمدى الربيعي

2-7 العلاقات بين مقاييس الترعة المركزية (الوسط

الحسابي والوسيط والمنوال)

2-8 المدى

2-9 الانحراف المتوسط

2-10 التباين والانحراف المعياري

2-11 معامل الاختلاف أو التغير

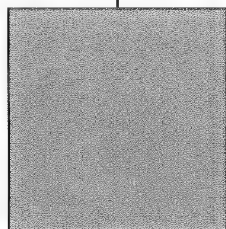
2-12 الدرجة المعيارية

2-13 الغصن والورقة

2-14 الرسم الصندوقي

القطر الأسود

2



الفصل الثاني

مقاييس النزعة المركزية والتشتت

Measures of Central Tendency and Dispersion

2-1 مقدمة Introduction

إن معظم القيم لمختلف الظواهر الطبيعية تتركز عادة في الوسط أو قريبة منه، ومقاييس التمرکز أو التوسط لأي مجموعة من البيانات التابعة للظاهرة هي تلك المقاييس التي تبحث في تقدير قيمة تتمركز حولها أغلبية البيانات بحيث تمثلها أفضل تمثيل وهناك ثلاث مقاييس إحصائية مهمة وشائعة الاستخدام، ويمكن أن تستخدم لتمثيل البيانات الإحصائية وحسب نوعية البيانات، فيمكن أن يكون الوسط الحسابي Arithmetic Mean هو من أهم المقاييس الإحصائية لكونه يستخدم جميع البيانات الإحصائية، أما المقياس الثاني فهو الوسيط Median ويعتبر من المقاييس المهمة والمستخدم بشكل واسع جداً وخاصة عندما يكون قسم من البيانات كبيرة جداً أو صغيرة جداً، أو ما تسمى في الإحصاء بالقيم الشاذة Outlier Values، أما المقياس الثالث فهو المنوال Mode والذي يستخدم بشكل واسع ويعتبر أيضاً من المقاييس المهمة وخاصة عند البيانات الوصفية وتستخدم التكرارات الإحصائية. كما وسيتم في هذا الفصل التعرف على المدى، الانحراف المتوسط، التباين والانحراف المعياري كمقاييس للتشتت والتي تقيس التغيرات الموجودة في البيانات والتي بدورها مع مقاييس النزعة المركزية تعطي صورة واضحة عن البيانات. وسيتم عرض جميع هذه المقاييس كالآتي:

2-2 الوسط الحسابي The Arithmetic Mean

الوسط الحسابي هو أحد أهم مقاييس النزعة المركزية وأكثرها استخداماً وشيوعاً وغالباً وعادة ما يسمى المعدل average ويفضل على جميع مقاييس النزعة المركزية لكونه يستعمل جميع البيانات ويستخدم الصيغ الرياضية.

2-2-1 الوسط الحسابي للبيانات الخام Raw Data أو ما يسمى بالقيم غير المبوبة:

الوسط الحسابي للبيانات الخام هو مجموع القيم مقسوماً على عددها. ولهذا ففي حالة وجود القيم المتطرفة الكبيرة جداً أو الصغيرة جداً أو ما تسمى القيم الشاذة (outlier values) فيفضل استخدام الوسيط. ويمكن أن يحسب الوسط الحسابي كما يلي :

الوسط الحسابي لمجموعة n من المشاهدات أو القيم X_1, X_2, \dots, X_n هو \bar{X} ويمكن إيجاده

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum X_i}{n}$$

للقيم من $I = 1, 2, \dots, n$

مثال
(1)

أوجد الوسط الحسابي للبيانات التالية :

63 , 66 , 67 , 68 , 69 , 70 , 71 , 72 , 74 , 75

الحل

$$\bar{X} = \frac{\sum X_i}{n} = \frac{63 + 66 + 67 + \dots + 75}{10} = \frac{695}{10} = 69.5$$

2-2-2 الوسط الحسابي للبيانات المجمعة Grouped Data أو ما يسمى بالقيم المبوبة:

عندما تكون البيانات موضوعة على شكل جدول توزيع تكراري ، سوف يتم استخدام مركز الفئة X_i والذي يساوي مجموع الحدين الأدنى والأعلى مقسوماً على 2 لتمثيل تلك الفئة. وعندما تتوزع البيانات توزيعاً طبيعياً فإن جميع المقاييس الوسط الحسابي والوسيط والمنوال لها نفس القيمة. أما عند رسم المنحني التكراري ويكون الطرف الأيسر أطول من الطرف الأيمن ، أي أطول نحو القيم الصغيرة فإن قيمة الوسيط تكون أكبر من قيمة الوسط الحسابي أما إذا كان الطرف الأيمن أطول من

الطرف الأيسر أي أطول نحو القيم الكبيرة فإن قيمة الوسط الحسابي تكون أكبر من الوسيط.

مثال (2)

الجدول التكراري التالي يمثل أطوال 40 حبة من الفول قيست إلى أقرب سم. المطلوب حساب الوسط الحسابي للطول؟

طول حبة الفول (x_i)	عدد الحبات (f_i)	$x_i f_i$
6	2	12
11	4	44
16	7	112
21	14	294
26	8	208
31	5	155
$\sum x_i f_i = 825$		$\sum x_i f_i = 825$

بتكملة الجدول أعلاه لإيجاد العمود $x_i f_i$ والذي يمثل حواصل ضرب مراكز الفئات والتكرارات نستطيع إيجاد الوسط الحسابي \bar{X} باستخدام الصيغة:

$$\bar{X} = \frac{\sum X_i f_i}{\sum f_i} = \frac{825}{40} = 20.6$$

ولهذا فإن الوسط الحسابي لطول حبة الفول هو 20,6 سم، أي أن معدل طول الحبة هو 20,6 سم.

مثال (3)

الجدول التكراري التالي يمثل تصنيف عدد من الأسر حسب إنفاقهم الشهري بالدينار.

المطلوب: حساب الوسط الحسابي للأجور.

فئات الإنفاق الشهري	عدد الأسر (f_i)	x_i	$x_i f_i$
0-	2	50	100
100-	5	150	750
200-	10	250	2500
300-	3	350	1050
400-500	1	450	450
	$\sum f_i = 21$		$\sum x_i f_i = 4850$

الحل

بتكملة الجدول أعلاه لإيجاد العمود x_i والعمود $x_i f_i$ نجد أن:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{4850}{21} = 230.95$$

ولهذا فإن معدل إنفاق الأسرة الشهري هو 230,95 ديناراً ويمكن القول بأنه 231 ديناراً.

2-2-3 خصائص الوسط الحسابي

من أهم خصائص الوسط الحسابي \bar{X} ما يلي:

1- مجموع انحرافات القيم عن وسطها الحسابي تساوي صفراً، أي أن:

أولاً- في حالة البيانات غير المبوبة:

$$\sum (x_i - \bar{X}) = 0$$

ثانياً- في حالة البيانات المبوبة:

$$\sum f_i (x_i - \bar{X}) = 0$$

البرهان:

$$\begin{aligned}
 \sum (X_i - \bar{X}) &= \sum X_i - \sum \bar{X} \\
 &= \sum X_i - n\bar{X} \\
 &= \sum X_i - \sum X_i \\
 &= 0 \\
 \sum f_i(X_i - \bar{X}) &= \sum f_i X_i - \bar{X} \sum f_i \\
 &= \sum f_i X_i - \left(\frac{\sum f_i X_i}{\sum f_i} \right) \sum f_i \\
 &= \sum f_i X_i - \sum f_i X_i \\
 &= 0
 \end{aligned}$$

مثال
(7)

لتوضيح الخاصية في حالة البيانات غير المبوبة. سيتم إثبات أن مجموع انحرافات القيم 2, 3, 5, 6, 4 عن وسطها الحسابي يساوي صفراً بالاستعانة بالجدول التالي:

(X_i)	$(X_i - \bar{X})$
4	0
5	1
6	2
3	-1
2	-2
$\bar{X} = 4 \quad \sum X_i = 20$	$\sum (X_i - \bar{X}) = 0$

مثال (5)

لتوضيح الخاصية في حالة البيانات المبوبة كما في الجدول التالي:

مراكز (x_i)	التكرار (f_i)	($x_i f_i$)	($X_i - \bar{X}$)	($X_i - \bar{X}$) f_i
3	3	9	-6.5	-19.5
6	5	30	-3.5	-17.5
9	10	90	-0.5	-5.0
12	8	96	2.5	20
15	4	60	5.5	22
	30	285	$\bar{X} = 285 / 30 = 9.5$	Zero

وهنا كذلك نبرهن أن مجموع انحرافات القيم عن وسطها الحسابي تساوي صفراً في حالة البيانات المبوبة أيضاً.

2- مجموع مربعات انحرافات القيم عن وسطها الحسابي اقل ما يمكن، أي اقل من مجموع مربعات انحرافات القيم عن أية قيمة غير الوسط الحسابي، أي أن:

$$\sum (X_i - \bar{X})^2 < \sum (X_i - B)^2$$

حيث أن B أي قيمة أخرى غير \bar{X}

البرهان:

نفرض ان B هو قيمة تمثل وسط فرضي غير الوسط الحسابي و المطلوب ان نجد

$$\sum (X_i - B)^2 \text{ هي اكبر من قيمة } \sum (X_i - \bar{X})^2 :$$

$$\sum (X_i - B)^2 = \sum (X_i^2 - 2BX_i + B^2)$$

$$= \sum X_i^2 - 2B \sum X_i + \sum B^2$$

$$= \sum X_i^2 - 2nB\bar{X} + nB^2$$

وبإضافة وطرح $n(\bar{X})^2$ من أعلاه ينتج:

$$\begin{aligned} &= \sum Xi^2 - 2nB\bar{X} + nB^2 + n(\bar{X})^2 - n(\bar{X})^2 \\ &= \left(\sum Xi^2 - n(\bar{X})^2 \right) + n(B^2 - 2B\bar{X} + (\bar{X})^2) \\ &= \sum (Xi - \bar{X})^2 + n(B - \bar{X})^2 \end{aligned}$$

وهذا يعني ان مجموع مربعات الانحرافات عن أي قيمة غير الوسط الحسابي هي اكبر من مجموع مربعات الانحرافات عن الوسط الحسابي وقدره $n(B - \bar{X})^2$ وهذه القيمة موجبة لكون الكمية تربيع ويمكن توضيح هذه الخاصية بالمثال التالي:

مثال (6)

بالرجوع للبيانات في مثال 1، والذي تم فيه استخراج قيمة \bar{X} تكون 4، نستطيع اثبات الخاصية الثانية كالاتي:

$$Xi : 4 \quad 5 \quad 6 \quad 3 \quad 2$$

$$\sum (Xi - \bar{X})^2 = (4-4)^2 + (5-4)^2 + (6-4)^2 + (3-4)^2 + (2-4)^2 = 10$$

أما لو طرحنا قيمة غير الوسط الحسابي ولتكن : $B = 5$

فسوف يكون مجموع مربعات الانحرافات كما يلي:

$$\begin{aligned} \sum (Xi - B)^2 &= \sum (Xi - B)^2 \\ &= (4-5)^2 + (5-5)^2 + (6-5)^2 + (3-5)^2 + (2-5)^2 \\ &= 1^2 + 0 + 1^2 + (-2)^2 + (-3)^2 \\ &= 1 + 0 + 1 + 4 + 9 \\ &= 15 \end{aligned}$$

وهذا يعني ان 15 اكبر من 10 مجموع مربع الانحرافات عن الوسط الحسابي وان الفرق بينهما يساوي التالي:

$$15 - 10 = 5$$

$$n(B - \bar{X})^2$$

$$5(5 - 4)^2 = 5$$

والتي تساوي

3 - عند اضافة كمية ثابتة او طرح كمية ثابتة ولتكن C إلى أو من كل قيم من القيم فسوف يكون الوسط الحسابي الجديد هو الوسط الحسابي القديم مضافا اليه أو مطروحا منه الكمية الثابتة C، أي أن اذا افترضنا أن القيم القديمة هي X_i وأن القيم الجديدة هي Y_i فذلك يعني أن:

$$Y_i = X_i + C$$

وهذه الخاصية تعطي أن:

$$\bar{Y} = \bar{X} + C$$

البرهان:

$$Y_i = X_i + C$$

$$\sum Y_i = \sum (X_i + C)$$

$$\sum Y_i = \sum X_i + nC$$

$$\frac{\sum Y_i}{n} = \frac{\sum X_i}{n} + \frac{nC}{n}$$

$$\bar{Y} = \bar{X} + C$$

مثال
(7)

يمكن تطبيق الخاصية على البيانات للمثال السابق

$$X_i = 4, 5, 6, 3, 2,$$

الحل الوسط الحسابي

$$\bar{X} = \frac{\sum X_i}{n} = \frac{20}{5} = 4$$

إذا تم إضافة القيمة الثابتة (C=5) لجميع القيم فإن القيم الجديدة سوف تكون:

$$Y_i = 9, 10, 11, 8, 7$$

لذلك فإن الوسط الحسابي الجديد هو:

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{45}{5} = 9$$

وهذا يعني أنه عند إضافة القيمة 5 نحصل على أن:

$$\bar{Y} = \bar{X} + 5$$

$$\bar{Y} = 4 + 5 = 9$$

أما عند طرح كمية ثابتة من كل قيم العينة بالشكل $Y_i = X_i - C$ فنحصل على أن:

$$\bar{Y} = \bar{X} - C$$

ولو طرحنا الكمية (2) (C = 2) فإن المشاهدات تكون $Y_i = 2, 3, 1, 4, 0$ وإن الوسط الحسابي الجديد سوف يكون كما يلي:

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{10}{5} = 2$$

4- عند ضرب كل قيمة من قيم العينة في كمية ثابتة (C) فإن:

الوسط الحسابي للقيم الجديدة هو الوسط الحسابي للعينة أي القيم القديمة مضروبة في الكمية الثابتة (C)، أي أن بافتراض أن X_i هي قيم العينة (القيم القديمة) وأن Y_i هي القيم الجديدة والتي نتجت عن ضرب القيم X_i بالقيمة الثابتة C فإن:

$$Y_i = CX_i$$

وبالتالي فإن:

$$\bar{Y} = C\bar{X}$$

البرهان:

$$Y_i = CX_i$$

$$\sum Y_i = C \sum X_i$$

$$\frac{\sum Y_i}{n} = C \frac{\sum X_i}{n}$$

$$\bar{Y} = C\bar{X}$$

مثال (8)

المثال التالي لتوضيح الخاصية بالأرقام وبالرجوع لنفس البيانات السابقة:

$$X_i = 4, 5, 6, 3, 2$$

$$\bar{X} = 4 \quad \text{الـ دال} \quad \text{الوسط الحسابي}$$

فإذا تم ضرب القيم بالقيمة الثابتة 4 نحصل على $Y_i = 4X_i$ وستكون القيم الجديدة كما يلي:

$$Y_i : 16, 20, 24, 12, 8$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{80}{5} = 16 \quad \text{فهو } \bar{Y} \text{ الجديد}$$

والذي يساوي ضرب الوسط \bar{X} في الكمية الثابتة 4، حيث أن:

$$\bar{Y} = \bar{X} \cdot 4 = 4 \cdot 4 = 16$$

5 - الوسط الحسابي لمجموع عدة عينات هو مجموع الأوساط الحسابية لهذه العينات:

فلو فرضنا لدينا ثلاث عينات X, Y, Z

$$C_i = Z_i + Y_i + X_i \quad \text{فإن مجموع العينات هو}$$

$$\bar{C} = \bar{Z} + \bar{Y} + \bar{X} \quad \text{وان الوسط الحسابي لهذه العينات هو}$$

البرهان:

$$C_i = Z_i + Y_i + X_i$$

$$\sum C_i = \sum (Z_i + Y_i + X_i)$$

$$\sum C_i = \sum Z_i + \sum Y_i + \sum X_i$$

$$\frac{\sum C_i}{n} = \frac{\sum Z_i}{n} + \frac{\sum Y_i}{n} + \frac{\sum X_i}{n}$$

$$\bar{C} = \bar{Z} + \bar{Y} + \bar{X}$$

مثال (9)

المثال التالي لتوضيح الخاصية : افرض لدينا ثلاث عينات و المطلوب إثبات ان الوسط الحسابي لهذه العينات مجتمعة هو مجموع الأوساط الحسابية لهذه العينات الثلاث وكما في الجدول التالي:

Z_i	Y_i	X_i	$C_i = Z_i + Y_i + X_i$
4	5	7	16
3	6	2	11
5	4	8	17
2	3	12	17
6	12	6	24
$\bar{X} = 4$	$\bar{X} = 6$	$\bar{X} = 7$	$C = 17$

وهذا يعني أن الوسط الحسابي الجديد هو:

$$\bar{C} = \bar{Z} + \bar{Y} + \bar{X}$$

$$\bar{C} = 4 + 6 + 7 = 17$$

6- الوسط الحسابي الموزون The weighted mean:

إذا كان لكل قيمة من المشاهدات X_i وزن خاص بها يتناسب مع أهميتها ويرمز له W_i فيكون الوسط الحسابي في هذه الحالة بالشكل:

$$\bar{X} = \frac{\sum X_i W_i}{\sum W_i}$$

مثال (10)

أوجد المعدل لدرجات أحد الطلبة عند تخرجه من الجامعة و كان عدد المواد 7 وكذلك الساعات لكل مادة.

لحساب الوسط الحسابي \bar{X} نستخدم صيغة الوسط الحسابي الموزون وذلك لأن كل مادة لها عدد من الساعات التي تمثل الأوزان w_i وبذلك نستخدم الجدول التالي:

المواد	الدرجات X_i	الساعات الأوزان w_i	$X_i w_i$
1	60	2	120
2	70	3	210
3	75	4	300
4	80	3	240
5	85	3	255
6	65	2	130
7	90	3	270
	525	$\sum w_i = 20$	$\sum w_i x_i = 1525$

∴ الوسط الحسابي الموزون:

$$\bar{X} = \frac{\sum X_i w_i}{\sum w_i} = \frac{1525}{20} = 76.25$$

أما لو تم حساب الوسط الحسابي بدون أوزان هو:

$$\bar{X} = \frac{525}{7} = 75$$

2-3 الوسيط The Median

هو المقياس الثاني من مقاييس النزعة المركزية في الأهمية، ويحسب إذا تم ترتيب البيانات حسب حجمها تصاعدياً أو تنازلياً. والوسيط يكون القيمة التي تقع في وسط البيانات وهذه الميزة من ناحية جيدة كونه لا يتأثر بالقيم الكبيرة أو الصغيرة أو ما تسمى بالقيم الشاذة وهذه الخاصية تميزه على الوسط الحسابي ومن ناحية أخرى فهذه الخاصية تعتبر المأخذ على الوسيط هو كونه يستخدم هذه القيمة فقط ويهمل جميع المعلومات في القيم الباقية وهنا يفضل عليه الوسط الحسابي إذا لم تكن هناك قيم شاذة.

2-3-1 الوسيط للبيانات الخام Raw Data او البيانات غير المبوبة:

لحساب الوسيط بالنسبة للبيانات الخام، نقوم بترتيب البيانات من أصغر قيمة الى أكبر قيمة فإذا كان عدد المشاهدات n فردي، يكون الوسيط هو القيمة الوسطية، أما إذا كان عدد المشاهدات n زوجي، فيكون الوسيط هو متوسط القيمتين الوسطيتين.

مثال (11)

افرض كان لديك البيانات في العينة التالية والتي تمثل 7 مشاهدات ، احسب الوسيط ؟

2 , 6 , 20 , 5 , 4 , 7 , 5

الحل

ترتب البيانات تصاعدياً من أصغر قيمة إلى أكبر قيمة

القيم : 2 4 5 5 6 7 20

الترتيب: 1 2 3 4 5 6 7

القيمة التي تسلسلها The Value $Q_2 = \frac{(n+1)}{2}$ تمثل موقع الوسيط

$$Q_2 = \frac{(7+1)}{2} = 4$$

إذا القيمة التي تقابل تسلسل 4 ، أي أن القيمة الرابعة وهي 5 تمثل الوسيط

مثال (12)

افرض كان لديك البيانات في العينة التالية والتي تمثل 8 مشاهدات ، احسب الوسيط ؟

4 , 7 , 5 , 8 , 9 , 7 , 10 , 9

الحل

ترتب البيانات تصاعدياً من أصغر قيمة الى أكبر قيمة

القيم: 4 5 7 7 8 9 9 10
الترتيب: 1 2 3 4 5 6 7 8

$$Q2 = \frac{(8+1)}{2} 4.5$$

بما أن عدد المشاهدات زوجي فإننا نحتاج لحساب الوسيط أن نحسب الوسيط الحسابي للقيمتين الوسيطيتين واللتين هما الرابعة والخامسة لكون موقع الوسيط هو (5, 4).

$$\text{The Median } M = \frac{(7+8)}{2} = 7.5$$

2-3-2 الوسيط للبيانات المجمعة Grouped Data او البيانات المبوبة:

عندما تكون البيانات موجودة في جدول توزيع تكراري فان الفئة التي تقابل رتبة الوسيط والتي تساوي نصف مجموع التكرارات في جدول التوزيع التكراري المتجمع الصاعد ، تسمى فئة الوسيط (Median Class) أو الفئة الوسيطة ويمكن إيجاد قيمة الوسيط من خلال فئة الوسيط هذه باستعمال القانون التالي:

$$M = L + \left[\frac{\frac{\sum f_i}{2} - F}{f} \right] W$$

حيث أن:

L تمثل الحد الأدنى الحقيقي لفئة الوسيط

W تمثل عرض فئة الوسيط

$\sum f_i$ تمثل مجموع التكرارات الكلية

f تمثل تكرار فئة الوسيط

F تمثل التكرار المتجمع الصاعد الذي يقابل الحد الأدنى الحقيقي لفئة الوسيط

مثال (13)

أوجد الوسيط لجدول التوزيع التكراري التالي والذي يمثل درجات 001 طالب من طلبة جامعة عمان الأهلية في مادة الإحصاء، فيما يلي جدول التوزيع التكراري مع التوزيع التكراري المتجمع الصاعد والنازل:

الترددات	(f)	الحدود المتكاملة للترددات	$F_1 \uparrow$	الحدود المتكاملة للترددات	$F_1 \downarrow$
60-	5	less than 63	5	60 & more	100
63-	18	less than 61	23	63 & more	95
66-	42	less than 69	65	66 & more	77
69-	27	less than 72	92	69 & more	35
72-75	8	less than 75	100	72 & more	8

الحل

أولاً نوجد رتبة الوسيط والذي يمثل عدد المشاهدات مقسوم على 2 ، ولهذا فإن $100 / 2 = 50$ تمثل رتبة الوسيط، إذا رتبنا البيانات تصاعدياً أو تنازلياً ومن خلال جدول التوزيع التكراري المتجمع فإن ترتيب القيمة 50 يقع بين التكرارات المتجمعة 65، 23 لذلك فإن قيمة الوسيط تقع ضمن الفئة {66 ، 69} وهذه الفئة تسمى فئة الوسيط، ولهذا فالوسيط هو حسب القانون سيكون كما يلي:

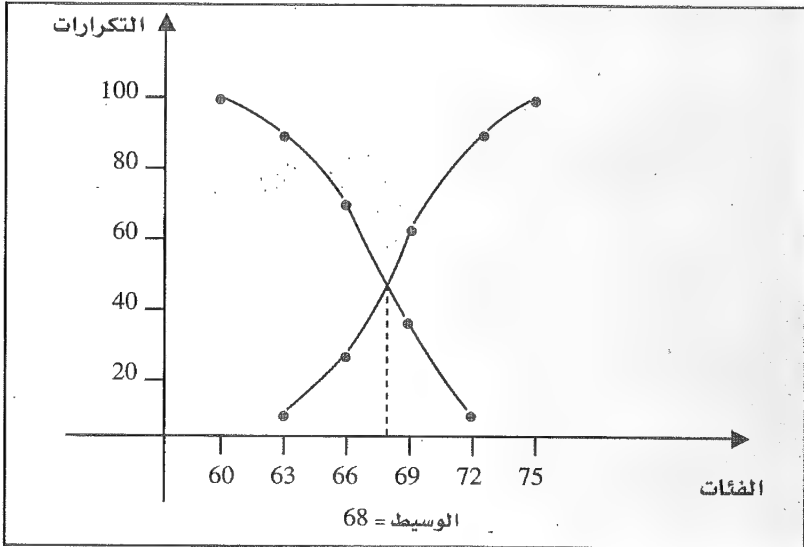
$$M = 66 + \left[\frac{\frac{100}{2} - 23}{42} \right] 3$$

$$M = 66 + \left[\frac{50 - 23}{42} \right] 3$$

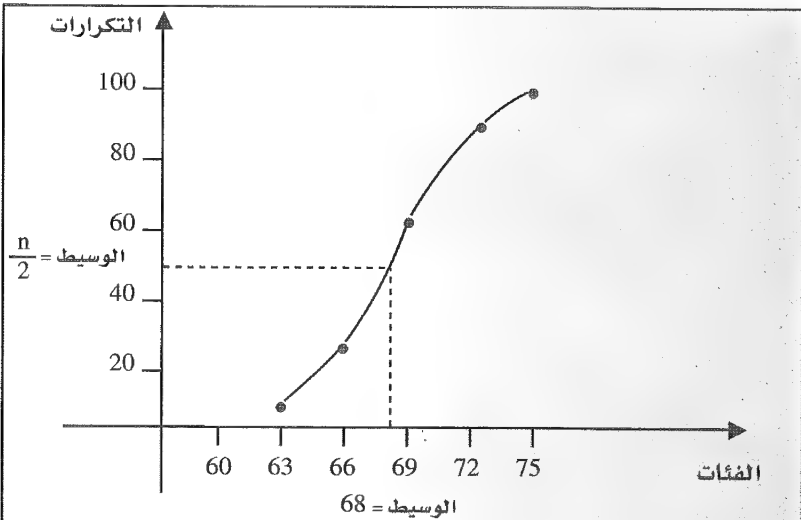
$$M = 66 + 1.929 = 67.929$$

أما لإيجاد الوسيط بطريقة الرسم فيمكن ملاحظة واستخدام المنحنين الصاعد والنازل حيث أن نقطة تقاطعهما تمثل إحداثي سيني بقيمة الوسيط وإحداثي صادي بقيمة

رتبة الوسيط كما يظهر بالشكل (1) التالي. أو يمكن ملاحظة واستخدام المنحنى الصاعد فقط لتحديد قيمة الوسيط والذي يمثل الإحداثي السيني لنقطة بحيث أن إحداثيها الصادي هو رتبة الوسيط $(\sum f_i / 2 = n / 2)$ كما يظهر في الشكل (2) التالي:



الشكل (1) الوسيط باستخدام المنحنيين الصاعد والنازل



الشكل (2) الوسيط باستخدام المنحنى الصاعد

2-4 المنوال The Mode:

المنوال هو المقياس الثالث من مقاييس النزعة المركزية في الأهمية والاستخدام وهو القيمة التي تكرر أكثر من غيرها، المنوال يفضل على غيره كونه سهل الحساب ، ولا يوجد تأثير للقيم الشاذة عليه أيضا، ولكن بصوره عامة لا يستخدم بالتحليلات اللاحقة وذلك لطريقة حسابه من التكرارات لكونه غالبا ما يستخدم المتغيرات الاسمية مثلا الجنس نوع السيارة او البلدان المنتجة إلى السيارات وهنا تستخدم التكرارات ولهذا لا يستخدم بشكل واسع مثل باقي المقاييس السابقة الوسط والوسيط.

2-4-1 المنوال للبيانات الخام Raw Data أو البيانات غير المبوبة:

مثال
(14)

أوجد المنوال لكل من المجاميع التالية :

- (a) 5 , 5 , 7 , 5 , 4 , 6 , 5 , -9 , 2 , 1 , 5 , 5 , 4
(b) 9 , 6 , 4 , 3 , 12 , 19 , 8 , 1 , 16 , 20 , 10 , 17
(c) 5 , 6 , 6 , 5 , 2 , 8 , 5 , 3 , 2 , 2

الحل حسب التسلسل هو

- (i) المنوال هو رقم 5 (القيمة التي تكررت أكثر من غيرها)
(ii) لا يوجد منوال.
(iii) هناك قيمتين للمنوال وهما 5 و 2 ، ويسمى التوزيع لمثل هذه الحالة bimodal والتي تعني أكثر من منوال.

2-4-2 المنوال للبيانات المجمعة Grouped Data أو البيانات المبوبة

عندما توضع البيانات على شكل جدول توزيع تكراري، وترتب على شكل فئات، فالفئة التي تملك أكبر تكرار تسمى الفئة المنوالية Modal Class. ويمكن تقدير المنوال وحسابه من خلال الفئة المنوالية باعتباره مركز تلك الفئة المنوالية كقيمة تقريبية للمنوال وكذلك يمكن استخدام القانون التالي:

$$M_0 = L + \left(\frac{d_1}{d_1 + d_2} \right) W$$

حيث أن:

L تمثل الحد الأدنى الحقيقي لفئة المنوال

W تمثل عرض فئة المنوال

d_1 الفرق بين تكرار الفئة المنوالية وتكرار الفئة السابقة لها

d_2 الفرق بين تكرار الفئة المنوالية وتكرار الفئة اللاحقة لها

كذلك يمكن استخراج قيمة المنوال باستخدام رسم المدرج التكراري (Histogram) وسيتم توضيح ذلك في المثال التالي:

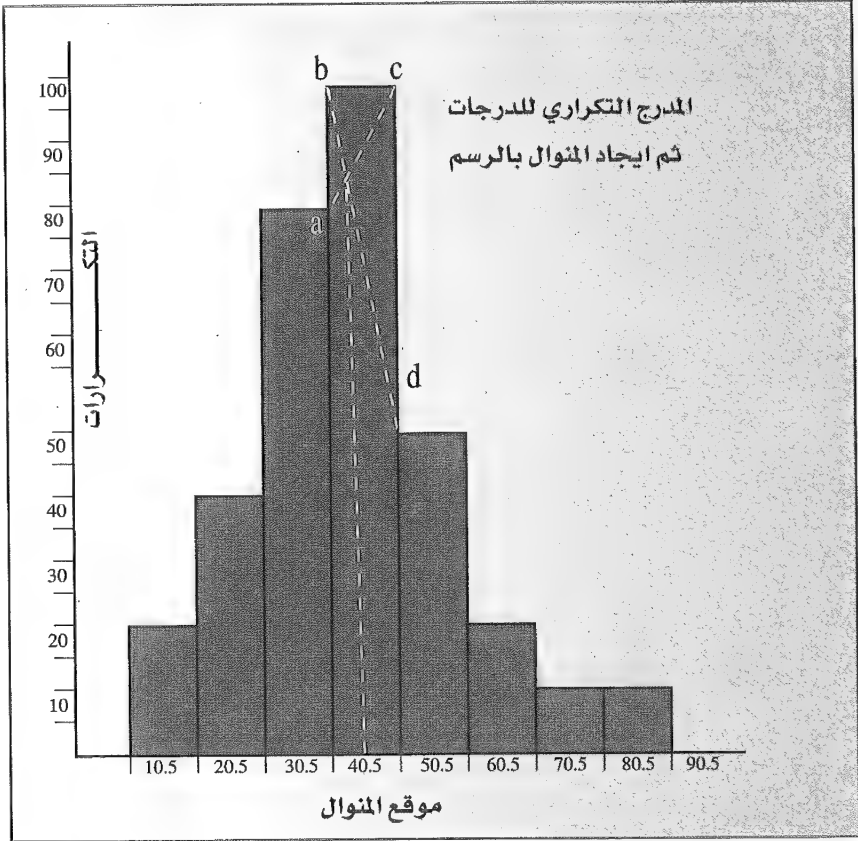
مثال (15)

أوجد المنوال لجدول التوزيع التكراري التالي، والذي يمثل درجات 330 طالب في إحدى الامتحانات العامة؟

التكرارات النسب	11-20	21-40	31-80	41-100	51-50	61-20	71-10	81-10	91-101
	20	40	80	100	50	20	10	10	0

الحل

هناك طريقتين لإيجاد المنوال وهما أولاً: بطريقة الرسم من خلال المدرج التكراري (Histogram) كما يتضح بالشكل (3) التالي:



الشكل (3) المنوال بطريقة الرسم

من خلال جدول التوزيع التكراري أو حتى من خلال الرسم للمدرج نستطيع تحديد الفئة المنوالية والتي هي كما يلي:

الفئة المنوالية هي (50 و 40) . وذلك لكونها تقابل أكبر تكرار.

ويمكن تقدير المنوال من خلال المدرج التكراري من خلال رسم خطين مستقيمين متقاطعين أحدهما يربط الحد الأعلى (a) للفئة قبل المنوالية بالحد الأعلى للفئة المنوالية (c) والآخر يربط الحد الأدنى للفئة المنوالية (b) بالحد الأدنى للفئة اللاحقة للفئة المنوالية (d) ومن نقطة تقاطعهما ننزل عمود على الاحداثي السيني الذي يعطي أن المنوال يساوي 43 تقريبا وذلك واضح من خلال الرسم.

أما الطريقة الثانية فتتم باستخدام قانون الفئة المتوالية التالي وكما يلي :

$$M_0 = L + \left(\frac{d_1}{d_1 + d_2} \right) W$$

$$M_0 = 41 + \left(\frac{20}{20 + 50} \right) 10$$

$$M_0 = 41 + \frac{20}{70} \times 10$$

$$M_0 = 43.6$$

لتطبيقات SPSS:

يمكن إيجاد جميع مقاييس النزعة المركزية السابقة كما يلي:

أن نختار Analyze من القائمة الرئيسية ومنها نختار Descriptive statistics
ومن هذا الخيار نختار Frequencies وبعد النقر على هذا الخيار نحدد المتغير المطلوب
ونختار Statistics من هذه الشاشة ومن هذا الخيار نحدد (mean, median, mode,...
ثم نقر على continue وبعدها ok.

2-5 الوسط الهندسي The Geometric Mean:

2-5-1 الوسط الهندسي للبيانات الخام او البيانات غير المبوبة

إذا كان لدينا n من المشاهدات

$X_1, X_2, X_3, \dots, X_n$

فإن الوسط الهندسي والذي يرمز له بالرمز هو:

$$\bar{G} = \sqrt[n]{X_1 * X_2 * \dots * X_n} \dots \dots \dots (1)$$

ولتسهيل إيجاد الوسط الهندسي نقوم بأخذ اللوغاريتم لطرفي المعادلة (1)

فتكون كما يلي:

$$\log \bar{G} = \left(\frac{1}{n} \right) \log (X_1 * X_2 * \dots * X_n)$$

$$\log \bar{G} = \frac{\log X_1 + \log X_2 + \dots + \log X_n}{n}$$

مثال
(16)

وهذا يوضح أن لو غارتم الوسط الهندسي لمجموعة من القيم هو الوسط الحسابي للوغاريتمات هذه القيم. ومن ثم نوجد معكوس اللوغاريتم لايجاد الوسط الحسابي الهندسي. والمثال التالي لتوضيح هذه الصيغة.

الحل

أوجد الوسط الهندسي للبيانات التالية: 5, 10, 8, 9, 7

$$\bar{G} = \sqrt[5]{5 \times 10 \times 8 \times 9 \times 7}$$

$$\log \bar{G} = \frac{\log 5 + \log 10 + \log 8 + \log 9 + \log 7}{5}$$

$$\log \bar{G} = \frac{0.69897 + 1 + 0.9031 + 0.954 + 0.8451}{5} = \frac{4.4012}{5}$$

$$\log \bar{G} = 0.88$$

$$\bar{G} = 7.59$$

ولكون حساب الوسط الهندسي يحتاج إلى اللوغاريتم فلهذا يجب أن تكون القيم في الوسط الهندسي جميعها موجبة وغالبا ما يستخدم الوسط الهندسي في معدلات التغير في المبيعات أو في الأرقام القياسية للأسعار.

2-5-2 الوسط الهندسي للبيانات المجمعة أو البيانات المبوبة:

لو فرضنا أن $X_1, X_2, X_3, \dots, X_n$ تمثل مراكز الفئات في جدول التوزيع التكراري والتكرارات التالية مقابلة لها على التوالي $f_1, f_2, f_3 \dots f_n$

فإن:

$$\bar{G} = \sqrt[n]{X_1^{f_1} \times X_2^{f_2} \times X_3^{f_3} \times \dots \times X_n^{f_n}}$$

وباستخدام اللوغاريتمات يكون الوسط الهندسي هو:

$$\log \bar{G} = \frac{\sum f_i \log X_i}{\sum f_i}$$

$$\log \bar{G} = \frac{f_1 \log X_1 + f_2 \log X_2 + \dots + f_n \log X_n}{\sum f_i}$$

2-6 الربعيات، والعشيرات (المئينيات) والهدى الربيعي:

Quartiles, Deciles, Percentiles, Interquartile Range

مقاييس النزعة المركزية السابقة لها استخداماتها الخاصة بها وهذه المقاييس الربعيات والعشيرات والمئينيات لها استخداماتها، فمثلاً يستخدم الوسيط لتحديد القيمة التي تقع في منتصف البيانات أو تقسم المساحة تحت المنحني التكراري إلى قسمين متساوية. ويمكن أن تبرز الحاجة إلى تقسيمات أخرى للبيانات أو تقسيمات المساحة تحت المنحني ولهذا فالربعيات تقسم البيانات إلى أربعة أقسام متساوية أو تقسم المساحة تحت المنحني إلى أربعة أقسام متساوية وكذلك الحال بالنسبة للمئينات تقسم المساحة أو البيانات إلى مائة جزء متساوية وهذه التقسيمات لها استخدامات واسعة وكثيرة جداً فمثلاً يمكن تقسيم مجموعة من الطلبة حسب المعدل إلى أربعة أقسام متساوية أو تقسيم المجموعة نفسها إلى مائة قسم ويمكن تحديد موقع كل طالب في هذه المجموعة حسب التقسيمات السابقة أو تقسيم مجموعة من الأشخاص حسب الدخل إلى أربعة أقسام أو مائة قسم متساوية وحسب التسلسل ويمكن تحديد أي ترتيب لأي من الأشخاص فالربع الثاني Q2 هو نفسه العشير الخامس D5 وكذلك المئين الخمسون P50.

ويمكن تعريف هذه التقسيمات كما يلي:

القيم الثلاث التي تقسم توزيع البيانات إلى أربعة أجزاء متساوية تعرف على أنها الربعيات Quartiles.

القيم التسعة والتسعون التي تقسم توزيع البيانات إلى مائة قسم متساوية تعرف بالمئينات Percentiles.

القيم التسعة التي تقسم توزيع البيانات إلى عشرة أقسام متساوية تسمى بالعشريات أو العشريات Deciles.

افرض أن هناك n من الأرقام مرتبة ترتيب تصاعدي، فيمكن إيجاد الربعيات والمئينيات كما يلي:

1) Lower Quartile $Q1 = \frac{1}{4} (n+1)$ th Value.

موقع قيمة الربع الأول ($Q1$) تقابل ترتيب البيانات $(n+1)/4$

2) Median $Q2 = \frac{1}{2} (n+1)$ th Value.

موقع قيمة الربع الثاني ($Q2$) وهو الوسيط تقابل ترتيب البيانات $(n+1)/2$

3) Upper Quartile $Q3 = \frac{3}{4} (n+1)$ th Value.

موقع قيمة الربع الثالث ($Q3$) تقابل ترتيب البيانات $3(n+1)/4$

4) 5th Decile $D5 = 5(n+1)/10$ th value

موقع قيمة العشير الخامس (الوسيط) تقابل ترتيب البيانات $7(n+1)/10$

5) 7th Decile $D7 = 7(n+1)/10$ th value

موقع قيمة العشير السابع تقابل ترتيب البيانات $10(n+1)/10$

6) 10th Percentile $P10 = 10(n+1)/100$ th Value.

موقع قيمة المئين العاشر ($P10$) تقابل ترتيب البيانات $(n+1)/100$

7) 90th Percentile $P90 = 90(n+1)/100$ th Value.

موقع قيمة المئين التسعون ($P90$) تقابل ترتيب البيانات $90(n+1)/100$

وهكذا إلى قيمة أي مئين (And So On)

8) The Interquartile Range = Upper quartile - Lower quartile

= $Q3 - Q1$

المدى الربيعي هو الفرق بين الربع الثالث ($Q3$) والربع الأول ($Q1$)

9) The Semi Interquartile Rang = $\frac{1}{2} (Q3 - Q1)$

نصف المدى الربيعي هو نصف الفرق بين الربع الثالث ($Q3$) والربع الأول ($Q1$)

10) The 10 to 90 Percentile Range = P90 - P10

المدى المئيني هو الفرق بين المئين التسعون والمئين العاشر.

ويلاحظ أن: الفائدة من تلك المئينات هو أنها تعتمد بشكل كامل على النصف الأوسط من البيانات ولهذا لا تتأثر بالقيم المتطرفة الكبيرة والصغيرة.

تطبيقات :

يمكن إيجاد جميع هذه التقسيمات للبيانات كما يلي:

أن نختار Analyze من القائمة الرئيسية ومن هذه القائمة نختار Descriptive statistics ومن هذه القائمة نحدد Frequencies وهنا نحدد المتغير ومن خيار Statistics نحدد Quartiles وكذلك Percentile ونحدد المطلوب ونقر على .ok

مثال (17)

أوجد نصف المدى الربيعي للبيانات التالية:

2 , 3 , 3 , 9 , 6 , 12 , 11 , 8 , 2 , 3 , 5 , 7 , 5 , 4 , 4 , 5 , 12 , 9 , 6

الحل نرتب البيانات تصاعدياً

2 , 2 , 3 , 3 , 3 , 4 , 4 , 5 , 5 , 5 , 6 , 6 , 7 , 8 , 9 , 9 , 11 , 12 , 12

هناك 19 رقماً

$Q1$ is the $1/4 (19 + 1)$ value = 5 (the fifth value)

وبذلك فإن قيمة الربع الأول القيمة الخامسة وبهذا ستكون $Q1 = 3$

$Q3$ is the $3/4 (19 + 1)$ value = 15 (the fifteenth value)

أي أن قيمة الربع الثالث القيمة الخامسة عشر وبهذا ستكون $Q3 = 9$

لهذا ستكون قيمة نصف المدى الربيعي هو:

$$\text{Semi inter-Range} = \frac{1}{2} (Q3 - Q1)$$

$$= \frac{1}{2} (9 - 3) = 3$$

وأخيراً فإن قيمة نصف المدى الربيعي هو 3.

أما لإيجاد المئينات باستخدام التوزيعات التكرارية فيفضل استخدام طريقة الرسم كما تم عمله لإيجاد الوسيط بالرسم، حيث أن رسم المنحنى المتجمع الصاعد يعطينا فكرة عن تسلسل البيانات وبالتالي معرفة مواقع بعض القيم المهمة مثل الوسيط، الربع الأدنى، الربع الأعلى، المئين العشرين، المئين التسعيني وغيرهم. وسيتم توضيح ذلك بالمثال التالي:

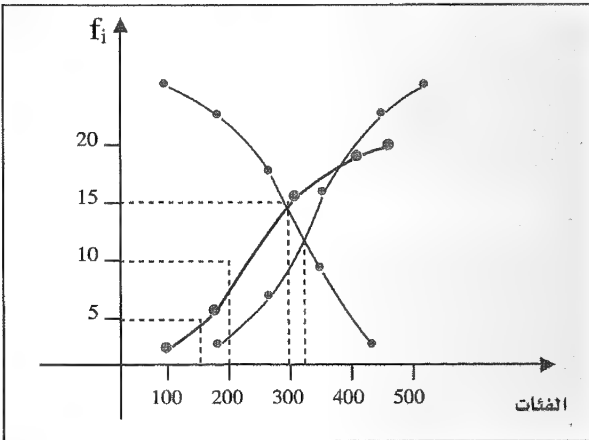
مثال (18)

استخدم بيانات الجدول التكراري التالي لتحديد قيمة Q_1 ، Q_2 ، Q_3 :

البيانات	(f_i)	الحدود الفعلية للبيانات	$F_i \uparrow$
0-	2	less than 100	2
100-	5	less than 200	7
200-	9	less than 300	16
300-	3	less than 400	19
400-500	1	less than 500	20
20			

الحل

باتباع خطوات رسم المنحنى المتجمع الصاعد جد أن الرسم يظهر كما في الشكل



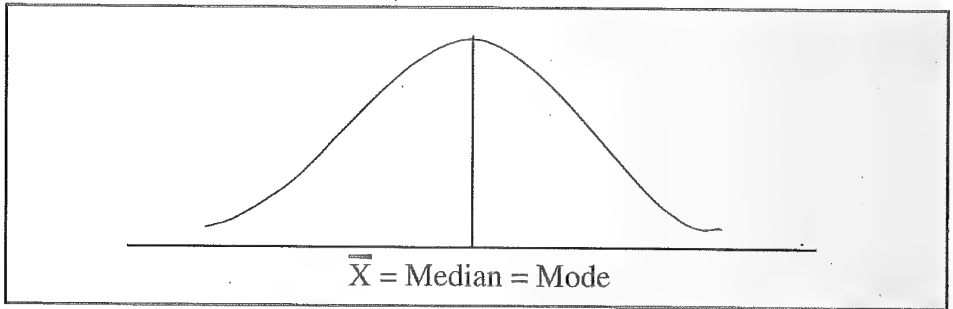
(5) التالي. وتحديد
رتب Q_3 ، Q_2 ، Q_1
على أنها 5، 10، 15
على التوالي نجد أن
قيم Q_3 ، Q_2 ، Q_1
هي 250، 170،
290 على التوالي:

الشكل (4) إيجاد Q_3 ، Q_2 ، Q_1 من المنحنى المتجمع الصاعد

2-7 العلاقة بين مقاييس النزعة المركزية (الوسط الحسابي والوسيط والمنوال):

بعد دراسة مقاييس النزعة المركزية بصورة عامة والمقاييس الوسط الحسابي والوسيط والمنوال بصورة خاصة لأهميتهم، لا بد من الدخول في معرفة وتحديد العلاقة بينهم، وسيتم توضيح ذلك كالآتي:

1- إذا كان التوزيع متماثلاً فإن قيمة كل من الوسط الحسابي \bar{X} والوسيط Median والمنوال Mode متساوية وكما تفهم بالشكل (5) التالي:



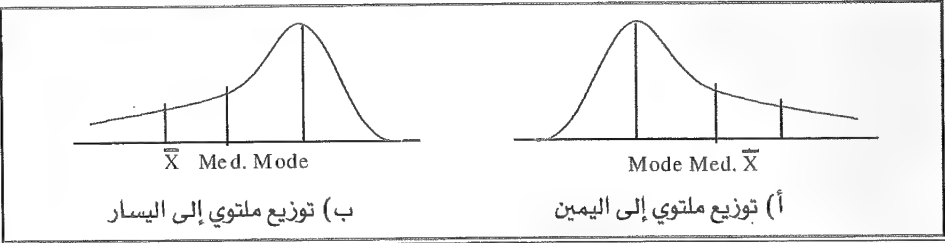
الشكل (5) العلاقة بين المقاييس للتوزيعات المتماثلة

ولكن يجب الملاحظة هنا بأن هذه التوزيعات الكاملة التماثل هي بصورة عامة توزيعات نظرية كالتوزيع الطبيعي Normal. بمعنى أن في الحياة العملية قد لا يكون هذا التماثل كاملاً وإنما يقال بأن التوزيع قريباً من التماثل وهنا تصبح المقاييس الثلاثة تقريباً متساوية، بمعنى معرفتنا لاتنين منهما يكفي للحصول على قيمة تقريبية للثالث بمعرفة أن التوزيع قريباً من التماثل وباستخدام العلاقة التقريبية التالية:

$$\bar{X} - \text{Mode} = 3 (\bar{X} - \text{Median})$$

وسيتم توضيح ذلك بالأمثلة لاحقاً.

2- أما إذا كان التوزيع مائلاً فيمكن أن يكون حسب الشكل (6) التالي:



الشكل (6) العلاقة بين المقاييس للتوزيعات الملتوية

ويلاحظ من الشكل (6) أعلاه أن قيم المقاييس مختلفة نوعاً ما إضافة إلى أن هناك ترتيب في قيمهم حسب درجة الالتواء.

ويجب الذكر هنا بأن قياس درجة الالتواء والذي يسمى معامل الالتواء يصبح مهماً في مثل هذه التوزيعات وأن هذا المعامل يتم حسابه بتعريف العزوم للتوزيعات التكرارية وأن هذا المعامل يعتمد على العزم الثالث $\sum (x_i - \bar{x})^3$ ولا بد من الذكر هنا بأن معامل الالتواء للتوزيع الطبيعي مساوياً للصفر وذلك لعدم وجود التواء فيه.

مثال (19)

الجدول التكراري التالي يمثل عدد من العمال مصنّفين حسب أجورهم الشهري. أوجد \bar{X} ، Mode، Median ثم ارسم التوزيع وحدد مواقع المقاييس عليه ومن ثم حدد شكله

فئات الأجور الشهرية	عدد العمال	x_i	$(x_i \cdot f_i)$	$F_i \uparrow$
100-	5	105	525	5
110-	12	115	1380	17
120-	20	125	2500	37
130-	10	135	1350	47
140-150	3	145	435	50
	50		6190	

يتضح لنا بعد إكمال الحسابات الخاصة بالجدول ومنها x_i ، $x_i f_i$ و F_i يمكن إيجاد الوسط الحسابي والوسيط والمنوال كالآتي:

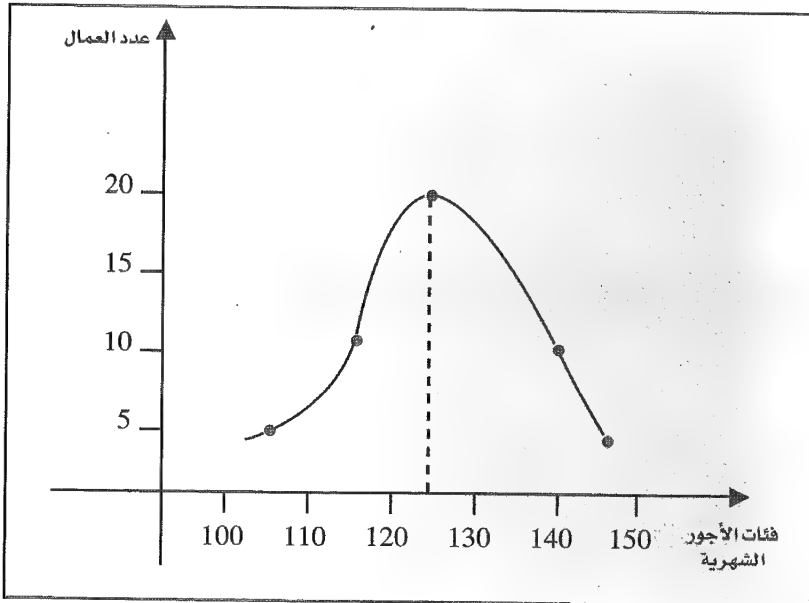
$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{6190}{50} = 123.8 \approx 124$$

$$Med. = L + \left(\frac{\frac{n}{2} - F}{f} \right) W$$

$$Med. = 120 + \left(\frac{25 - 17}{20} \right) 10 = 124$$

$$Mode = \frac{120 + 130}{2} = 125$$

وبما أن القيم متقاربة جداً من بعضها، يمكن اعتبارها متساوية ويلاحظ ذلك بشكل أفضل من الشكل (7) التالي:



الشكل (7) مواقع المقاييس على التوزيع

2-8 المدى The Range:

هو الفرق بين أعلى قيمة و أصغر قيمة، ويعتمد بشكل كامل على القيمتين المتطرفتين.

ولكون المدى يعتمد على هاتين القيمتين لذلك فإنه يتأثر بهذه القيم بشكل كبير جداً وخاصة في حالة كون إحدى القيم أو الاثنتين قيم شاذة وهذا المأخذ على المدى قلل من أهميته وكذلك من استخدامه مع سهولته الكبيرة.

مثال (20)

أوجد المدى للبيانات التالية:

10, 15, 19, 34, 90, 50, 70, 56, 85, 58, 57, 20, 86, 85, 87,
66, 67, 77, 80, 84, 81, 86, 13

الحل

نحدد أولاً أكبر قيمة وهي 90 وأصغر قيمة وهي 10 ليكون المدى كالاتي:

$$\text{The Range} = 90 - 10 = 80$$

مثال (21)

أوجد المدى الربيعي للمثال (20) أعلاه.

الحل

أولاً نرتب البيانات تصاعدياً ثم نوجد الربيعيات بعد تحديد مواقعها وكما يلي:

2		6																				
10	13	15	19	20	34	50	56	57	58	66	67	70	77	80								
81	84	85	85	86	86	87	90															
18					22																	

$$Q1 = \frac{1}{4} * (n+1) \text{ value} = 6 \text{ value} \quad \text{ونجد أن}$$

$$Q1 = 34 \quad \text{أي أن}$$

$$Q3 = \frac{3}{4} * (n+1) \text{ value} = 18 \text{ value} \quad \text{أي أن}$$

$$Q3 = 85 \quad \text{لذلك فإن}$$

$$\text{Interquartile Range} = Q3 - Q1$$

$$= 85 - 34$$

$$= 51$$

(c) المدى المئيني هو القيمة التي يمكن إيجادها من خلال الفرق بين المئينات
P10, P90 (Percentiles)

أوجد المدى المئيني للمثال (1) السابق ؟

$$P10 = 10(n+1) / 100 \text{ value} = 240/100 \text{ value}$$

$$= 2.4 \text{ value} = 2.00 \text{ value}$$

$$P90 = 90(n+1) / 100 \text{ value} = 2160/100 \text{ value}$$

$$= 21.6 \text{ value} = 22.00 \text{ value}$$

$$P10 = 13$$

$$P90 = 87$$

$$P10 \text{ to } P90 \text{ Range} = P90 - P10 \quad \text{أذن المدى المئيني يساوي}$$

$$= 87 - 13$$

$$= 74$$

2-9 الانحراف المتوسط The Mean Deviation

2-9-1 الانحراف المتوسط للبيانات الخام:

ويعرف على أنه متوسط الانحرافات المطلقة عن الوسط الحسابي ويستخدم جميع البيانات :

الانحراف المتوسط لمجموعة من البيانات عددها n

يمكن ان تحسب بواسطة القانون التالي :

$$Meandivision = \frac{\sum |Xi - \bar{X}|}{n}$$

$$i = 1, 2, 3, \dots, n$$

حيث أن هو الوسط الحسابي للمجموعة وأن $|Xi - \bar{X}|$ هي القيمة المطلقة (قيمة موجبة) للفروقات

مثال
(22)

أوجد الانحراف المتوسط للبيانات التالية :

196 198 199 200 200 201 201 202 205 198

الحل

نوجد أولا الوسط الحسابي ليكون:

$$\bar{X} = \frac{\sum Xi}{n} = \frac{2000}{10} = 200$$

Xi	$Xi - 200$	$ Xi - 200 $
196	-4	4
198	-2	2
198	-2	2
199	-1	1
200	0	0
200	0	0
201	1	1
201	1	1
202	2	2
205	5	5
Total		18

$$Meandivision = \frac{\sum |Xi - \bar{X}|}{n} = \frac{18}{10} = 1.8$$

لذلك فإن الانحراف المتوسط يساوي 1.8 .

2-9-2 الانحراف المتوسط للبيانات المجمعة:

$$Meandivision = \frac{\sum |X_i - \bar{X}| f_i}{\sum f_i}$$

$$i = 1, 2, \dots, n$$

الانحراف المتوسط للبيانات المبوبة:

مثال
(23)

أوجد الانحراف المتوسط لجدول التوزيع التكراري التالي :

x_i	f_i	$x_i f_i$	$X_i - \bar{X}$	$ X_i - \bar{X} $	$ X_i - \bar{X} f_i$
3	4	12	-7.57	7.57	30.28
6	6	36	-4.57	4.57	27.42
9	10	90	-1.57	1.57	15.70
12	12	144	1.43	1.43	17.16
15	6	90	4.43	4.43	26.58
18	4	72	7.43	7.43	29.72
	42	444			146.86

$$\bar{X} = \frac{\sum X_i f_i}{\sum f_i} = \frac{444}{42} = 10.57$$

$$Meandivision(M.D.) = \frac{\sum |X_i - \bar{X}| f_i}{\sum f_i}$$

$$M.D. = \frac{146.86}{42}$$

$$M.D. = 3.4966$$

2-10 التباين والانحراف المعياري Variance and standard deviation

كما ذكرنا في مقدمة هذا الفصل أن تعريف التشتت أو الاختلاف على أنه التقارب أو التباين بين المشاهدات داخل المجموعة وبالتالي فإن مقاييس التشتت وبصورة عامة، التباين والانحراف المعياري بصورة خاصة تقيس مدى تشتت البيانات أو المشاهدات عن وسطها الحسابي \bar{X} ، وكلما كان مقياس التشتت أكبر كلما دل ذلك على عدم تجانس المشاهدات.

ويعرف التباين على أنه مجموع مربعات انحرافات القيم x_i عن وسطها الحسابي \bar{X} مقسوماً على مقام مناسب هو $(n-1)$ والذي يسمى بدرجات الحرية degrees of freedom. وهذا المفهوم له أهمية كبيرة في دراسات التوزيعات وخاصة تلك المناسبة لحجوم العينات الصغيرة، كما سيتم توضيح ذلك في الفقرات القادمة من هذا الكتاب.

التباين إلى مجموعة من البيانات يعرف (S^2) وهو:

$$S^2 = \frac{\sum (x_i - \bar{X})^2}{n-1}$$

لذلك فإن الانحراف المعياري هو:

$$S = \sqrt{S^2}$$

2-10-1 التباين والانحراف المعياري للبيانات الخام:

الانحراف المعياري لمجموعة من البيانات تتكون من n من الأرقام

x_1, x_2, \dots, x_n مع الوسط الحسابي \bar{X} هو S .

$$S = \sqrt{\frac{\sum (x_i - \bar{X})^2}{n-1}}$$

الانحراف المعياري يعتبر من أهم مقاييس التشتت أو الانتشار، الوحدات بالنسبة للانحراف هي نفسها التي في البيانات الأصلية.

مثال (24)

أوجد الانحراف المعياري (S) للبيانات التالية :

196 198 198 199 200 200 201 201 202 205

الحل

من القيم نجد أن $\bar{X} = \frac{2000}{10} = 200$ وبذلك نستطيع تكوين الجدول التالي:

X	x-200	(x-200) ²
196	-4	16
198	-2	4
198	-2	4
199	-1	1
200	0	0
200	0	0
201	1	1
201	1	1
202	2	4
205	5	25
2000		56

ولذلك فإن S^2 هو $S^2 = \sum (X_i - 200)^2 / 9 = 56/9$

أي أن $S^2 = 6.22$

وبذلك فإن الانحراف المعياري هو: $S = \sqrt{6.22} = 2.49$

هناك صيغة أخرى لحساب الانحراف المعياري، لكون الصيغة السابقة في بعض

الأحيان صعبة التطبيق وخاصة عندما تكون قيمة الوسط الحسابي كسور بدلا من قيمة صحيحة.

$$S^2 = \frac{\sum Xi^2 - (\sum Xi)^2 / n}{n-1}$$

$$S = \sqrt{\frac{\sum Xi^2 - (\sum Xi)^2 / n}{n-1}}$$

مثال (25)

أوجد الوسط الحسابي والانحراف المعياري للأعداد التالية:-

2, 3, 5, 6, 8

$$\bar{X} = \frac{\sum X}{n} = \frac{24}{5} = 4.8 \text{ ليكن } \bar{X}$$

الانحراف المعياري بالطريقة الأولى كما يلي:

X	$Xi - \bar{X}$	$(Xi - \bar{X})^2$	Xi^2
2	-2.8	7.84	4
3	-1.8	3.24	9
5	0.2	0.04	25
6	1.2	1.44	36
8	3.2	10.24	64
24		22.8	138

$$S = \sqrt{\frac{\sum (Xi - \bar{X})^2}{n-1}}$$

$$S^2 = \frac{22.8}{4} = 5.7$$

$$S^2 = 5.7$$

$$S = 2.387$$

أما الطريقة الثانية لحساب الانحراف المعياري كما يلي:

$$S = \sqrt{\frac{\sum Xi^2 - (\sum X)^2/n}{n-1}}$$

$$S^2 = \frac{\sum Xi^2 - (\sum X)^2/n}{n-1}$$

$$S^2 = \frac{138 - (24)^2/5}{4}$$

$$S^2 = \frac{22.8}{4} = 5.7$$

$$S = 2.387$$

وهذا يعني أن الطريقة الثانية أبسط وأسهل في إيجاد الانحراف المعياري.

2-10-2 التباين والانحراف المعياري للبيانات المبوبة (جدول التوزيع التكراري):

افرض القيم التالية تمثل مراكز الفئات لجدول التوزيع التكراري (مراكز الفئات تستخدم لإيجاد الانحراف المعياري)

$$X_1, X_2, X_3, \dots, X_n$$

وأن القيم التالية هي التكرارات المقابلة لها:

$$f_1, f_2, f_3, \dots, f_n$$

لذلك فإن الانحراف المعياري نجده بالصيغة التالية:

$$S = \sqrt{\frac{\sum f_i(X_i - \bar{X})^2}{\sum f_i}}$$

$$i = 1, 2, 3, \dots, n$$

أما الصيغة الثانية البديلة فهي:

$$S = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i} - \bar{X}^2}$$

$$i = 1, 2, 3, \dots, n$$

مثال (26)

أوجد الانحراف المعياري والتباين لجدول التوزيع التكراري التالي:

(x_i) مراكز الفئات	(f_i) التكرار	$x_i f_i$	$(x_i - \bar{X})$	$(x_i - \bar{X})^2$	$(x_i - \bar{X})^2 f_i$
25	4	100	50-26.21	686.96	2747.84
36	6	216	-15.21	231.344	1388.064
47	10	470	-4.21	17.724	177.24
58	5	290	6.79	46.104	230.52
69	5	345	17.79	316.5	1582.5
80	4	320	28.79	828.86	3315.44
315	34	1741			9441.504

الحل

باستخدام الطريقة الأولى نجد أن الوسط الحسابي

$$\bar{X} = \frac{1741}{34} = 51.21$$

$$S^2 = \frac{\sum f(x_i - \bar{X})^2}{\sum f_i} = \frac{9441.504}{34} = 277.69 \text{ وأن التباين هو } 277.69$$

وأخيرا الانحراف المعياري هو: $S = \sqrt{277.69} = 16.663$

(x_i) مراكز الفئات	(f_i) التكرار	x_i^2	$x_i f_i$
25	4	625	2500
36	6	1296	7776
47	10	2209	22090
58	5	3364	16820
69	5	4761	23805
80	4	6400	25600
	34		98591

أما بالطريقة الثانية فنجد:

$$S^2 = \frac{\sum x_i^2 f_i}{\sum f_i} - \bar{X}^2$$

أي أن:

$$S^2 = \frac{98591}{34} - 2622.5$$

$$S^2 = 2899.74 - 2622.5$$

$$S^2 = 277.24$$

$$S = \sqrt{277.24}$$

$$S = 16.65$$

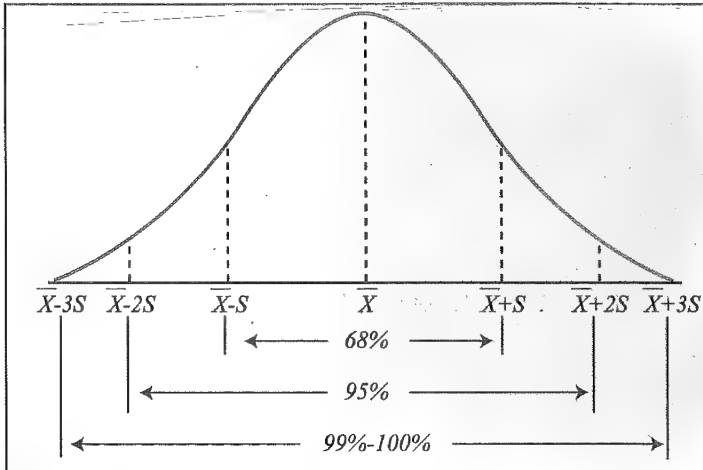
تطبيقات SPSS:

يمكن إيجاد جميع مقاييس التشتت كما يلي:

أن نختار Analyze من القائمة الرئيسية ومنها نحدد Descriptive statistics وبعد تحديد المتغير المطلوب نختار الخيار options ومن هذه الشاشة نحدد (mean; variance; std. devia; Range) ونقر على continue ثم ok.

2-10-3 تفسير الانحراف المعياري:

يستخدم الانحراف المعياري بصورة عامة كمقياس لتشتت العينة في محاولة معرفة عدد البيانات ونسبتها تلك التي تقع على بعد معين من الوسط الحسابي حيث أنه وباستخدام الصيغة التجريبية The empirical rule (أي في التوزيعات المتماثلة والطبيعية) نلاحظ أن نسب البيانات تكون كما هي موضحة بالشكل (8) التالي:



الشكل (8) الصيغة
التجريبية للتوزيع
المتماثل

بمعنى أنه في التوزيعات الطبيعية نتوقع أنه 68% من القيم تقع على بعد انحراف معياري واحد حول الوسط، وأنه 95% من القيم تقع على بعد انحرافين معياريين حول الوسط وأخيراً فإن حوالي 99% إلى 100% أي جميع القيم تقع على بعد ثلاث انحرافات معيارية حول الوسط.

أما في التوزيعات الغير متمائل فعموماً نتوقع الحصول على نسب مختلفة عن تلك. وبذلك فلأجل توضيح عملية استخدام الانحراف المعياري لتفسير عينة ما، علينا إيجاد الفترات $\bar{X} \pm s$ ، $\bar{X} \pm 2s$ ، $\bar{X} \pm 3s$ ثم تحديد عدد البيانات ثم نسبتها وبالتالي مقارنتها مع النسب أعلاه لتحديد ما إذا كان توزيع البيانات طبيعي والتشتت معتدل، وسيتم توضيح ذلك بالمثل التالي. أما لمقارنة عينتين فيمكن مقارنة النسب ببعضها والقول بعد ذلك أن تشتت إحدى العينتين أكبر أو أقل من العينة الأخرى.

مثال (27)

أوجد نسبة البيانات التي تقع على بعد انحراف معياري ثم انحرافين معياريين ثم ثلاث انحرافات معيارية حول الوسط وقارنها بالنسبة من الصيغة التجريبية للبيانات التالية:

10.5	13.5	9.5	8.2	6.5	8.4	8.1	6.9	7.5	13.5
9.6	7.2	7.1	9.0	9.9	8.2	13.2	9.2	6.9	7.7
10.6	9.7	7.5	7.2	5.9	6.6	11.1	8.8	5.2	8.2
9.4	11.3	5.6	10.1	8.0	8.5	11.7	7.1	7.7	6.0
6.8	8.0	7.4	10.5	7.8	7.9	6.5	6.9	6.5	9.5

الحل

بديهي أن النسب من الصيغة التجريبية للأبعاد أعلاه هي 100%، 59%، 86%، أما هذه النسب فيمكن حسابها للبيانات أعلاه فيتم بعد حساب s ، \bar{X} حيث أن:

$$\bar{X} = 8.49, s = 1.98$$

والآن سنجد الفترة $\bar{X} \pm s$ لتكون (6.51, 10.47)

وبحساب عدد القيم الواقعة ضمن هذه الفترة نجد أنه 43 ، النسبة فهي 68%.

الآن سنجد الفترة $\bar{X} \pm 2s$ لتكون (4.53 , 12.45)

أما القيم الواقعة ضمنها فهي 47 والنسبة هي 94%.

وأخيراً نجد الفترة $\bar{X} \pm 3s$ لتكون (2.55 , 14.43)

وعدد القيم الواقعة ضمنها هي 50 ، النسبة فهي 100%.

وبذلك نلاحظ أن هذه النسب المحسوبة متفقة مع النسب من الصيغة التجريبية مما يدل نوعاً ما على أن التوزيع طبيعياً.

2-11 معامل الاختلاف أو التغير Coefficient of Variation:

الانحراف المعياري وحده لا يكفي لإعطاء صورة واضحة عن التشتت داخل مجموعة من البيانات، ولهذا ربما يكون معامل الاختلاف أكثر ملاءمة لإعطاء صورة واضحة عن التشتت داخل مجموعة من البيانات، وبذلك فإن معامل الاختلاف يفضل على الانحراف المعياري وغيره من مقاييس التشتت لمقارنة تشتت البيانات بين عدة مجاميع من البيانات.

ويعرف معامل الاختلاف (للمجتمع والعينة على التوالي) كما يلي:-

وهو يعطي نسبة الانحراف المعياري إلى الوسط الحسابي. وبما أن معامل الاختلاف هو مقياس لقياس التغير النسبي على شكل نسبة مئوية، إذن معامل التغير ممكن أن يستخدم لمقارنة التشتت داخل عدة مجاميع من البيانات حتى في حالة أن وحدات القياس للمجاميع تكون مختلفة.

$$C.V = \frac{S}{\bar{X}} \times 100\% \quad \text{OR} \quad C.V = \frac{\sigma}{\mu} \times 100\%$$

مثال (28)

أوجد معامل التغير أو الاختلاف (C.V) لمجموعتين من درجات الطلبة في مادة الإحصاء.

المجموعة الأولى g1	المجموعة الثانية g2	
12	33	الوسط الحسابي
24	56	الانحراف المعياري

الحل

$$C.V_{group_1} = \frac{24}{12} \times 100\% = 200\%$$
 نجد معامل الاختلاف المجموعة الأولى

$$C.V_{group_2} = \frac{56}{33} \times 100\% = 170\%$$
 ونجد معامل الاختلاف المجموعة الثانية

وبالمقارنة نجد أن التشتت في المجموعة الأولى أكبر من التشتت في المجموعة الثانية.

2-12 الدرجة المعيارية Standardized Scores

نحتاج في بعض الأحيان إلى مقارنة مشاهدات مجاميع مختلفة، وهنا نحتاج لتحويل هذه المشاهدات إلى وحدات قياسية حتى نتمكن أن نقوم بالمقارنة بين المشاهدات وذلك باستخدام الوسط الحسابي والانحراف المعياري لكل مجموعة من مجاميع المشاهدات وكما يلي:

باستخدام القانون التالي الخاص بـ (Zi) التي تمثل القيمة أو الدرجة المعيارية نجد أن:

$$Z_i = \frac{X_i - \bar{X}}{S_i}$$

حيث يمكن إيجاد قيمة Z لكل مجموعة من المجاميع باستخدام الوسط الحسابي الخاص بالمجموعة وكذلك الانحراف المعياري الخاص بالمجموعة أيضا وهنا تصبح قيمة (Z) خالية من الوحدات ولهذا نستطيع القيام بالمقارنة بين قيم Z التي كل واحدة تمثل المجموعة الخاصة بها.

حيث أن: هي الدرجة المعيارية Z_i
هو الوسط الحسابي \bar{X}_i
الانحراف المعياري S_i

وعند اخذ مجموعة واحدة تتكون من عدة مشاهدات وتم تحويل مشاهداتها إلى وحدات معيارية (Z_i) وذلك يطرح الوسط الحسابي للمجموعة منها وتقسيمها على الانحراف المعياري للمجموعة فإن الوسط الحسابي لهذه المشاهدات المحولة إلى (Z_i) أي الوحدات المعيارية) سوف يكون (1) و أن الانحراف المعياري لها سوف يكون (0). أي أن:

$$Z_i \sim N(0, 1)$$

وهذا يعني أن Z_i تتوزع توزيع طبيعي بوسط حسابي يساوي صفرا وانحراف معياري يساوي واحد.

مثال (29)

لمقارنة درجات أحد الطلبة لمادتين مختلفتين فإذا كانت درجته في مادة الإحصاء (90) و أن الوسط الحسابي للطلبة معه في مادة الإحصاء يساوي (80) والانحراف المعياري لدرجات الطلبة في مادة الإحصاء أيضا يساوي (5) وكانت درجته في مادة المحاسبة تساوي (80) علما بأن الوسط الحسابي للطلبة في مادة المحاسبة كان (65) و أن الانحراف المعياري للطلبة في مادة المحاسبة أيضا يساوي (5) فما هي افضل درجة للمادتين للطلاب.

الحل

أولا عند القيام بالمقارنة الاعتيادية نجد أن درجة الإحصاء (90) افضل من درجة المحاسبة (80).

وعند القيام بتحويل هاتين الدرجتين إلى معيارية نجد أن:

$$Z_i = \frac{X_i - \bar{X}}{S_i}$$

$$Z_1 = \frac{90-80}{5} = 2 \quad \text{لذلك فإن القيمة المعيارية لدرجة الإحصاء هي:}$$

$$Z_2 = \frac{80-65}{5} = 3 \quad \text{أما بالنسبة لمادة المحاسبة فإن } Z_i \text{ هي:}$$

وهذا يعني أن درجته في مادة المحاسبة افضل من درجته في مادة الإحصاء علما بأن مادة الإحصاء 90 أعلى من درجة المحاسبة ولهذا فيجب تحويل أي مشاهدات إلى درجات معيارية عند المقارنة لإعطاء الصورة الصحيحة والواضحة عن المقارنة.

2-13 الغصن والورقة Stem and Leaf

من الطرق الحديثة لعرض البيانات الإحصائية طريقة تدعى الغصن والورقة، وهذا الأسلوب من العرض اسهل لتكونه من جدول التوزيع التكراري وكذلك من المدرج التكراري وبصورة عامة فهو يعرض معلومات اكثر.

فهو يعرض نفس معلومات المدرج التكراري histogram وكذلك يعرض معلومات جدول التوزيع التكراري بالاضافة إلى ذلك فهو يعرض الأرقام بشكلها الاعتيادي عند ربط كل قيمة بين الغصن والورقة الخاص بها لهذا يسمى بالشكل الهجين hybrid لأن معلوماته تمثل الرسم وكذلك الأرقام في الجدول في آن واحد كما هو موضح لاحقا.

لتوضيح الغصن والورقة لاحظ المثال التالي:

مثال
(30)

أوجد الغصن والورقة للمشاهدات التالية:

70	64	99	55	64	89	87	65	62	38	67	70	60	69	78	39
75	56	71	51	99	68	95	86	57	53	47	50	55	81	80	98
51	36	63	66	85	79	83	70								

لبناء أو تكوين الغصن والورقة والذي هو في نفس الوقت يوضح البيانات على شكل مجاميع تشبه جدول التوزيع التكراري وكذلك يعرضها على شكل رسم إحصائي يشبه المدرج التكراري نقوم بالخطوات التالية:

(1) نختار أرقام من البيانات على أنها الأرقام التي تكون أمام البيانات أو تقود البيانات Leading digits والتي تشمل الجزء الأول من الأرقام والتي تمثل العشرات وهذا ينتج الأرقام التالية (9, , 4, 3) نضعهم على شكل عمود كما في الشكل اللاحق الغصن Stem.

(2) بعدها نبدأ بالمرور على جميع الأرقام لكتابة الجزء الثاني من كل رقم مقابل الجزء الأول منه Final digit ونضعه إلى اليمين من الجزء الأول وهي الآحاد.

الرقم الأول في جدول رقم (1) السابق (70)، ولهذا نحتاج أن نضع (0) صفر على يمين الرقم (7)، بعدها نقرأ الأرقام بالجدول عمود بعد الآخر ولهذا الرقم الثاني هو (75) ولهذا نحتاج لوضع (5) إلى يمين الرقم (7)، ونستمر على هذه الطريقة.

ولهذا سوف يكون شكل الغصن والورقة لبيانات المال كما موجود في الشكل رقم 1 التالي. وكما موضح في الشكل رقم 1 الأرقام التي تتبعها الأرقام الأخرى على اليمين تدعى الأرقام البداية الغصن stem والأرقام النهائية أو المكملة تسمى الأوراق Leaves والشكل بشكل كامل يسمى Stem-and-Leaf أو الغصن والورقة.

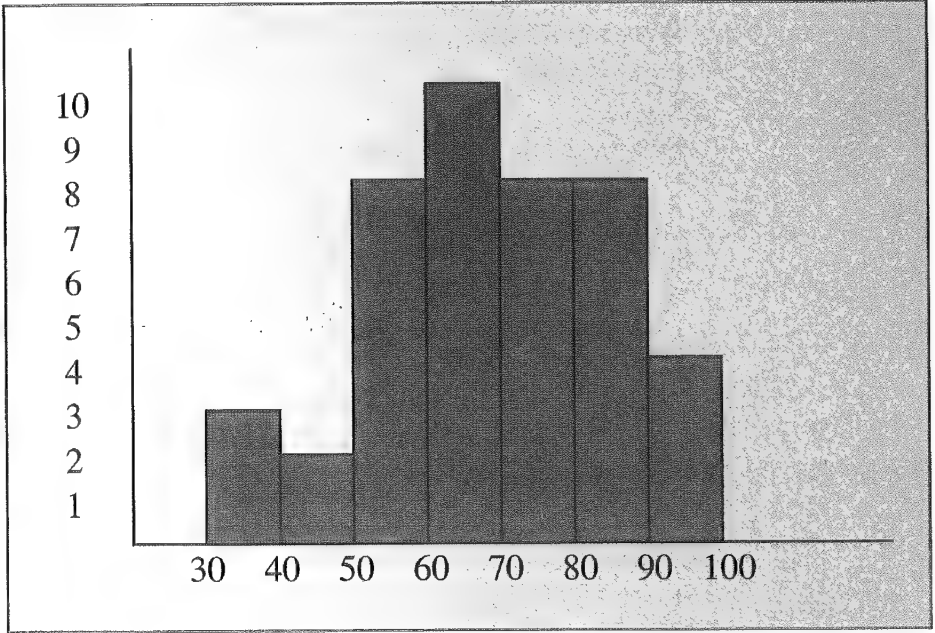
(a) (Stem and leaf)

الغصن Stem	الأوراق Leaves
3	8 6 9
4	7
5	7 1 6 3 5 1 0 5
6	2 4 7 3 6 4 0 9 8 5
7	0 5 1 0 9 8 0
8	5 9 1 7 0 3 6
9	9 9 5 8

جدول التوزيع التكراري للبيانات (b)

البيانات	تكرارات
30-39	3
40-49	1
50-59	8
60-69	10
70-79	7
80-89	7
90-99	4
Total	40

(c) Histogram



الشكل (9) الغصن والورقة وعلاقته مع المدرج التكراري

الشكل (9) (a) الغصن والورقة للبيانات السابقة هو مشابه إلى المدرج التكراري (b) لنفس البيانات لأن طول الأوراق في الغصن يساوي عدد التكرارات في كل فئة. لو قلبنا الغصن والورقة بدرجة 90° لمقارنة الغصن والورقة مع المدرج التكراري لوجدنا انهما متشابهان جدا.

نوع آخر من الغصن والورقة ويدعى شكل الغصن والورقة المرتب. لهذا النوع من الغصن والورقة الأوراق في كل صف من الصفوف ترتب من اصغر قيمة الى اكبر قيمة. والشكل الجديد يسهل من فهم البيانات وكذلك يسهل حساب المقاييس الإحصائية مثل مقياس الوسيط.

والشكل رقم 2 التالي يوضح هذا النوع أي الغصن والورقة المرتب على بيانات المثال (1) السابق وكما يلي:

Stem	Leaves
3	6 8 9
4	7
5	0 1 1 3 5 5 6 7
6	0 2 3 4 4 5 6 7 8 9
7	0 0 0 1 5 8 9
8	0 1 3 5 6 7 9
9	5 8 9 9

الشكل (10) الغصن والورقة المرتب

مثال (31)

أوجد شكل الغصن والورقة للبيانات التالية التي تمثل نسبة الكولسترول في الدم إلى 02 مريض.

210	209	212	208	202	218	200	214	218	210
217	207	210	203	215	221	213	210	199	208

الحل

بما أن الأعداد كل واحد يمثل ثلاثة أرقام سوف تستخدم أول رقمين كأغصان ونستخدم الرقم الثالث كأوراق. والغصن والورقة للبيانات كما يلي في الشكل رقم 3.

لكل غصن صف واحد (a)

Stem	Leaves
19	9
20	8 2 9 7 0 8 3
21	0 7 5 0 8 2 0 0 3 8 4
22	1

لكل غصن هناك صفين (b)

Stem	Leaves
19	
19	9
20	2 0 3
20	8 9 7 8
21	0 0 2 0 0 3 4
21	7 5 8 8
22	1
22	

الشكل (11) الغصن والورقة

الشكل (11) (a) وهو الغصن والورقة صف واحد لكل غصن يمكن ان يستفاد منه بشكل متوسط لكون عدد الغصون قليل نوعا ما والبيانات مجمعة بشكل كبير. أما الشكل (11) (b) فلكل غصن هناك صفين ولهذا نلاحظ البيانات بشكل اكثر وضوح لتوزيع البيانات وقد تم وضع الأوراق التي تقابل كل غصن في صفين من (0-4) في صف رقم واحد ومن (5-9) في صف رقم 2 وهذا النوع يفيد كثيرا عندما يكون حجم البيانات كبيرا حيث يتم ملاحظة توزيع البيانات بشكل واضح جدا.

تطبيقات SPSS:

يمكن إيجاد stem & leaf من SPSS من خلال الخطوات التالية:

أن نختار Analyze من القائمة الرئيسية ومن هذه القائمة نختار قائمة Descriptive statistics ومن هذه القائمة نختار الخيار Explore وبعد تحديد المتغير نختار من هذه الشاشة خيار plot ومن هذه الشاشة نحدد stem and leaf ونعدها نقر على continue ونعدها ok.

2-14 الرسم الصندوقي BoxPlot:

الرسم الصندوقي boxplot وأحيانا يسمى BoxPlot and Wisker Length وهذا النوع من الرسوم الإحصائية المهمة التي اكتشفت من قبل العالم J.W.Tukey (1977) ويعتبر من أقوى وأهم الرسوم الإحصائية الى الان منذ اكتشاف الرسوم الإحصائية وتأتي أهميته من حيث انه يعطي معلومات كاملة عن الخصائص المهمة للبيانات بعد ترتيبها تصاعديا فهو يوضح مركز البيانات 50% من منتصف البيانات والذي يسمى المدى الربيعي أي interquartile range ومختصرها (IQR) والذي يتم حسابه من الفرق بين الربع الثالث (Q3) والربع الأول (Q1) أي $(Q3 - Q1) = IQR$ والمدى الربيعي والذي يشمل الربع الثاني والربع الثالث من حجم البيانات أي منتصف البيانات من الوسط لأن الرسم الصندوقي Boxplot يقسم البيانات الى أربعة أرباع. ومن خلال الربعات الثلاث (Q1, Q2, Q3) نستطيع تحديد مركز البيانات وهو الوسيط ($median = Q2$) وكذلك نستطيع تحديد التغيرات بين الربع الثاني والربع

الأول هو بين الربع الثاني والأول للبيانات هو ($Q2 - Q1$) وكذلك التغير بين الربع الثاني والربع الثالث هو الفرق بين الربع الثالث والربع الثاني ($Q3 - Q2$) ولكن الربعيات الثلاث لا تعطي معلومات عن التغيرات داخل الربع الأول والربع الرابع ويمكن إيجاد التغيرات داخل الربع الأول من خلال الفرق بين الربع الأول وأصغر قيمة ($Q1 - \min$) والتغيرات داخل الربع الرابع هو من خلال الفرق بين أكبر قيمة والربع الثالث ($Q3 - \max$) وهذا يعني استخدام المعلومات من أصغر قيمة إلى أكبر قيمة وهذه الخاصية تعطي الأهمية إلى الرسم الصندوقي حيث يستخدم جميع البيانات ويستطيع التركيز أو التوضيح لأي جزء من الأجزاء.

أما الرسم الصندوقي المعدل modified box plot فيضع موقع القيم الشاذة outliers خارج جسم الرسم الصندوقي والتي يمكن تحديدها كما يلي:

تقع القيم خارج ما يلي $\{Lower\ limit = Q1 - 1.5 * IQR\}$ أي أقل من هذه القيمة، وكذلك القيمة التي تقع خارج الموقع التالي تعتبر قيمة شاذة $\{Upper\ limit = Q3 + 1.5 * IQR\}$. ويمكن توضيح ذلك من خلال الأمثلة اللاحقة لتوضيح الرسم الصندوقي والرسم الصندوقي المعدل. ومن خلال ما تقدم يمكن أن نقول أن من أقوى ومن أفضل طرق الرسم لعمل المقارنة بين عدة مجاميع من البيانات والذي يمكن عمل المقارنة على أساس مراكز البيانات (الوسيط The Median) أو على أساس الربع الأول أو الربع الرابع أو على أساس النصف الأوسط من البيانات 50% من وسط البيانات (IQR) المدى الربيعي للبيانات أو ممكن أن تكون المقارنة على أساس المجاميع بشكل كامل وحتى هل توجد هناك قيم شاذة صغيرة أو كبيرة. وهذه الميزة جعلت الرسم الصندوقي يكون أفضل طرق الرسم لهذه الأهمية. ولاحقاً سوف نلاحظ كيفية بناء كل نوع من الرسوم الصندوقية.

2-14-1 كيفية بناء الرسم الصندوقي البسيط Box Plot:

- 1 - نحدد القيم الخمسة للخلاصة وهي كما يلي ($\min, Q1, Q2, Q3, \max$).
- 2 - ارسم خط أفقي وحدد مكان الخمسة قيم على هذا الخط (الربعيات, $Q1, Q2, Q3$ وأكبر وأصغر قيمة).

أوجد الرسم الصندوقي البسيط للبيانات التالية:



عدد المشاهدات $N = 19$.

أولاً: نرتب البيانات تصاعدياً من أصغر قيمة إلى أكبر قيمة:

8 9 10 10 12 15 16 17 20 24 25 26 27 29 29

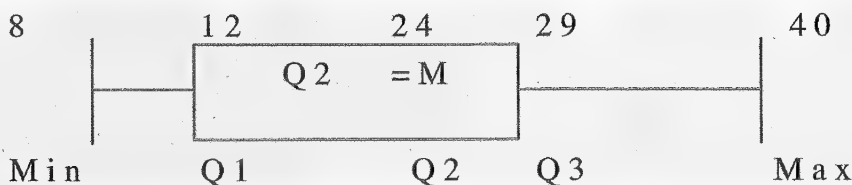
(1) الخمسة قيم الخلاصة مواقع (Q3, Q2, Q1). $Min = 8$

$$Q_1 = \frac{(n+1)}{4} \text{value} = \frac{20}{4} \text{value} = 5 \text{value}, Q_1 = 12$$

$$Q_2 = \frac{2(20)}{4} \text{ value} = \frac{40}{4} \text{ value} = 10 \text{ value}, Q_2 = 24$$

$$Q_3 = \frac{3(20)}{4} \text{ value} = \frac{60}{4} \text{ value} = 15 \text{ value}, Q_3 = 29$$

$$Max = 40$$



2-14-2 كيفية بناء الرسم الصندوقي المعدل modified boxplot:

1- لعمل الرسم الصندوقي المعدل للمثال 4 نحدد الربيعات (Q1, Q2, Q3) وهذه الربيعات للمثال هي:

$$(Q3 = 58 \quad Q2 = 34 \quad Q1 = 25)$$

2- نحدد القيم المجاورة adjacent values وكذلك القيم الشاذة وهي (minadjacent = 10, maxadjacent = 58) والقيمة الشاذة هي (outlier value = 105)

3- الآن نربط الربيعات على شكل مستطيل وثم ربط القيم المجاورة على شكل خطوط مستقيمة وهي:

(Min = 10 Max = 70) وهذه القيم تربط بمستطيل الرسم الصندوقي لكون المسافة عن المستطيل اقل من مرة ونصف من طول المدى الربيعي ($IQR \times 1.5$) في الاتجاهين من (Q1, Q3) أما القيم الشاذة فتوجد قيمة واحدة فقط ولهذا توضح بشكل منفصل وموقعها اكبر مرة ونصف من المدى الربيعي ($1.5 \times IQR$) من موقع الربيع الثالث (Q3) حيث ان:

{ $IQR = 58 - 25 = 33$ } وهنا تصبح المسافة من الربيع الثالث Q3 الى القيمة الشاذة اكبر من: ($1.5 \times IQR = 1.5 \times 33 = 49.5$)

$$Q3 + 49.5 = 58 + 49.5 = 107.5 \text{ ليكون}$$

ولهذا فالقيمة (108) تصبح قيمة شاذة لكونها ابعد من 5, 701 وتحدد بشكل منفصل.

مثال (33)

أوجد الرسم الصندوقي المعدل للبيانات التالية:

10 12 20 14 25 20 28 26 30 29 32 34 35 40 60
55 59 58 45 46 70 66 65

الحل

أولاً: نجد عدد المشاهدات $N = 23$.

ثانياً: نرتب البيانات تصاعدياً

الترتيب	القيم
1	10
2	12
3	14
4	20
5	20
6	25
7	26
8	28
9	29
10	30
11	32
12	34
13	35
14	40
15	45
16	46
17	55
18	58
19	59
20	60
21	56
22	66
23	70

ثالثاً: يجب إيجاد الخمسة قيم الخلاصة ومواقعها.

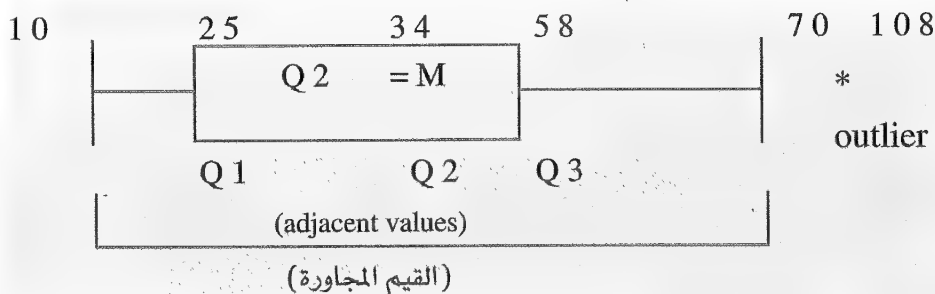
(Min Q1 Q2 Q3 Max)

$$Min = 10$$

$$\text{موقع } Q_1 = \frac{(n+1)}{4} = \frac{24}{4} = 6, Q_1 = 25$$

$$\text{موقع } Q_2 = \frac{2(n+1)}{4} = \frac{2(24)}{4} = 12, Q_2 = 34$$

$$\text{موقع } Q_3 = \frac{3(n+1)}{4} = \frac{3(24)}{4} = 18, Q_3 = 58$$



تسمى أكبر قيمة وأصغر قيمة التي تنهي الرسم الصندوقي بالقيم المجاورة عند تحديد القيم الشاذة، والقيم الشاذة تؤثر خارج جسم الرسم الصندوقي في حالة الرسم الصندوقي المعدل (modified boxplot).

2-14-3 استخدامات أخرى للرسم الصندوقية:

الرسم الصندوقية صممت بشكل أساسي لمقارنة مجموعتين أو أكثر من

البيانات. ولعمل هذه المقارنة يجب ان نرسم رسم صندوقي لكل مجموعة ونضع الرسومات على شكل ترتيب واحد باستخدام نفس وحدات القياس أما عمودياً أو أفقياً.

مثال (34)

استخدم الرسم الصندوقي لمقارنة مجموعتين من البيانات.



* أيضاً يمكن استخدام الرسوم الصندوقية لمقارنة أشكال توزيعات البيانات.

نظيقات SPSS:

يمكن إيجاد الرسم الصندوقي Boxplot كما يلي:

أن نختار Graphs من القائمة الرئيسة ومن هذه القائمة نختار الخيار Boxplot ومن هذه الشاشة نختار Simple ونحدد الخيار الثاني للبيانات Data وننقر Define ونحدد المتغير المطلوب لهذه الشاشة وننقر على الزر ok للتنفيذ.

أسئلة

الفصل الثاني

1- البيانات في جدول التوزيع التكراري التالي تمثل أعمار مجموعة من موظفي جامعة عمان الأهلية.

Classes الفئات	Frequency التكرار
18-23	10
24-29	14
30-35	20
36-41	16
42-47	8

المطلوب:

- التباين والانحراف المعياري.
- الانحراف المتوسط.
- الوسيط.
- معامل الاختلاف.

2- البيانات التالية تمثل درجات الحرارة لمجموعة من الدول لعدد من الأيام.

6	8	10	12	14	9	8	5		
-2	-4	0	-3	-7	3	2	1	5	
5	6	4	7	8	9	10	3	4	7

المطلوب:

- الوسط الحسابي.
- الوسيط.
- المنوال.
- المدى.
- التباين.
- الانحراف المعياري.
- الانحراف المتوسط.
- معامل الاختلاف.

3- اكمل الجدول التالي:

X	S^2	C.V
	20	10%
30	100	
10		15%

4- البيانات التالية تمثل درجات مجموعة من طلبة جامعة عمان الأهلية وتتكون من (72) طالباً.

80	90	85	25	30	40	45	46	60	55
65	85	86	77	65	60	66	67	78	53
80	10	50	56	59	62	64			

المطلوب:

- أوجد الرسم الصندوقي المعدل.
- أوجد الغصن والورقة.
- أوجد نصف المدى الربيعي.
- أوجد نصف المدى المثني.

5- البيانات التالية تمثل الأجور اليومية لمجموعة من العاملين:

16.49	7.49	13.78	16.70	7.44
10.40	10.75	5.51	17.92	17.21
13.10	14.70	10.68	15.27	19.72
14.02	15.99	15.07	15.55	16.67

المطلوب:

- الوسط الحسابي.
- الوسيط.
- المنوال.
- التباين والانحراف المعياري.

6- البيانات التالية تمثل قيم فواتير التلفون لمجموعة من الأسر:

47	41	28	15	08
10	40	30	32	11
38	37	42	12	05

المطلوب:

- الوسط الحسابي والوسيط والمنوال.
- الانحراف المعياري.

7- البيانات التالية تمثل تصنيف عدد من مرضى ضغط الدم المرتفع حسب مستوى الكوليسترول في الدم:

مستوى الكوليسترول	عدد المرضى
195-	1
200-	3
205-	4
210-	7
215-	4
220-225	1

المطلوب:

- الوسط الحسابي والوسيط والمنوال.
- الوسيط بطريقة الرسم.
- $Q1$ و $Q2$ و $Q3$ بطريقة الرسم.
- المنوال بطريقة الرسم.
- التباين والانحراف المعياري.
- نسبة البيانات التي تقع على بعد (1) انحراف معياري ثم (2) انحراف معياري ثم (3) انحراف معياري حول الوسط.

3

الفصل الثالث

الإنحدار والإرتباط الخطي البسيط

Simple Linear Regression and Correlation

3-1 مقدمة

3-2 معامل الارتباط الخطي البسيط

3-3 معامل الارتباط للرتب والصفات

3-4 الانحدار الخطي البسيط

3-5 العلاقة بين معاملات الانحدار ومعامل الارتباط

الفصل الثالث

3

الفصل الثالث

الانحدار والارتباط الخطي البسيط

Simple Linear Regression and Correlation

3-1 مقدمة Introduction

غالباً ما يكون من المهم دراسة العلاقة بين ظاهرتين أو أكثر وتحديد نوع وقوة تلك العلاقة كالعلاقة بين ظاهرتي الإنفاق والدخل أو العلاقة بين ظاهرتي زيادة الإنتاج والقوى العاملة أو العلاقة بين ظاهرتي الطول والوزن وغيرها. نظرية الارتباط عادة تظهر شدة أو قوة العلاقة بين الظاهرتين أو بين المتغيرين X و Y_i . أما دراسة هذه العلاقة من خلال التمثيل البياني بأفضل علاقة اقتران ممكنة بالشكل $Y_i = f(x)$ فتسمى بدراسة الانحدار ويسمى المستقيم أو المنحني الذي يمثل هذه الدالة بمستقيم أو منحني الانحدار.

الانحدار يعتبر أحد الأساليب الإحصائية المهمة والتي تستخدم بشكل واسع جداً ومنذ فترات طويلة لتحديد التأثيرات بين المتغيرات المستقلة X 's والمتغير المعتمد Y_i ويمكن أن توضع هذه المتغيرات على شكل معادلات خطية بحيث يمكن استخدامها للتنبؤ عن قيمة المتغير المعتمد Y_i بدلالة المتغيرات المستقلة X 's. فإذا كان المتغير Y_i يعتمد على متغير مستقل واحد X_i فنسمي الانحدار عندئذ الانحدار الخطي البسيط Simple Linear Regression، أما إذا كان المتغير Y_i يعتمد على عدد من المتغيرات المستقلة فيسمى الانحدار بالانحدار المتعدد Multiple Regression.

الانحدار والارتباط من الأساليب الإحصائية المهمة والواسعة الانتشار والاستخدام لعرض وإيجاد العلاقة بين متغيرين إذا كانت وحدات القياس للمتغيرات من النوع المستمر Continuous. مثلاً لدراسة السوق يمكن استخدام الانحدار والارتباط لإيجاد العلاقة بين الأموال المخصصة للدعاية لبضاعة معينة ومقدار أو كمية البضاعة المباعة أو الربح المقابل، ويمكن استخدام أيّاً من وسائل الدعاية فيمكن استخدام معامل الارتباط The Correlation Coefficient ويرمز له بالرمز r لقياس

العلاقة بين المتغيرين أو الظاهرتين. ويمكن أن نستخدم معادلة الانحدار Regression Equation لتقدير كمية البضاعة المباعة من خلال تحديد مقدار المبلغ المخصص للدعاية والأمثلة عديدة في جميع المجالات الاقتصادية وغيرها. وسوف نبدأ بمعامل الارتباط لتحديد العلاقة بين المتغيرين.

3-2 معامل الارتباط الخطي البسيط Correlation Coefficient:

المقياس الإحصائي المستخدم بشكل واسع لقياس العلاقة بين المتغيرين يسمى معامل الارتباط لبيرسون Person's Correlation Coefficient ويرمز له بالرمز r , حيث أن $-1 \leq r \leq 1$ فإذا كانت العلاقة قوية وموجبة (طردية) فإن قيمة r تقترب من 1 وإذا كانت قوية وسالبة (عكسية) فإن قيمة r تقترب من -1 وكلما اقتربت قيمة r من الصفر 0 فيعني ذلك أن العلاقة ضعيفة وهنا نستطيع القول كلما تقترب النقاط من خط الانحدار أو تقع على خط الانحدار فإن قيمة r تقترب من الواحد وكلما ابتعدت النقاط عن خط الانحدار فإن قيمة r تقترب من الصفر.

وهذا المقياس هو معامل بيرسون للارتباط يقيس قوة العلاقات الخطية فقط ولهذا ففي الحالات التي لا توجد فيها علاقات خطية بين المتغيرات أو الظواهر وإنما توجد علاقة غير خطية nonlinear relationship بين المتغيرين فعند ذلك يجب أن تعرض البيانات قبل استخدام مقياس الارتباط للتأكد من نوع العلاقة وكذلك لتحديد فيما إذا كانت هناك قيم شاذة Outlier values ويمكن قياس العلاقة بين المتغيرين X_i, Y_i إلى n من أزواج المشاهدات حسب قانون بيرسون Person's Correlation Coefficient التالي وهو أحد الصيغ الثلاثة:

$$r = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\left[\sqrt{\sum X_i^2 - n \bar{X}^2} \right] \left[\sqrt{\sum Y_i^2 - n \bar{Y}^2} \right]}$$

حيث أن:

\bar{X} يمثل الوسط الحسابي لقيم X

\bar{Y} يمثل الوسط الحسابي لقيم Y

والأمثلة التالية توضح نوع العلاقة الخطية بين المتغيرات.

مثال (1)

البيانات التالية تمثل الدخل الشهري والادخار لمجموعة من الأشخاص.
والمطلوب إيجاد معامل الارتباط لبيرون.

X_i : 10 12 11 7 6 8 9 5 4 3

Y_i : 4 5 5 3 2 3 3 1 1 0

الحل

X_i	Y_i	$X_i Y_i$	X_i^2	Y_i^2
10	4	40	100	16
12	5	60	144	25
11	5	55	121	25
7	3	21	49	9
6	2	12	36	4
8	3	24	64	9
9	3	27	81	9
5	1	5	25	1
4	1	4	16	1
3	0	0	9	0
75	27	248	645	99

حيث أن

$X_i Y_i$: هو حاصل ضرب القيم المتقابلة من X_i و Y_i .

X_i^2 : هو تربيع قيم X_i .

Y_i^2 : هو تربيع قيم Y_i .

\bar{X} : الوسط الحسابي لقيم X_i .

\bar{Y} : الوسط الحسابي لقيم Y_i .

$$\bar{X} = \frac{75}{10} = 7.5$$

$$\bar{Y} = \frac{27}{10} = 2.7$$

$$r = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\left[\sqrt{\sum X_i^2 - n \bar{X}^2} \right] \left[\sqrt{\sum Y_i^2 - n \bar{Y}^2} \right]}$$

$$r = \frac{248 - (10)(7.5)(2.7)}{\left[\sqrt{645 - 10(7.5)^2} \right] \left[\sqrt{99 - 10(2.7)^2} \right]} = 0.98$$

وهذا يعني وجود علاقة قوية وموجبة بين المتغيرين. فكلما زاد الدخل ازداد الادخار.

مثال (2)

البيانات التالية تمثل الادخار وحجم الأسرة لمجموعة من العوائل ذات الدخل المحدود أو المتساوي، والمطلوب إيجاد معامل الارتباط لبيرسون.

X_i : 3 5 10 9 8 7 4 2 6 1

Y_i : 6 4 0 1 2 3 5 6 4 7

X_i حجم الأسرة	Y_i الادخار	$X_i Y_i$	X_i^2	Y_i^2
36	9	18	6	3
16	25	20	4	5
0	100	0	0	10
1	81	9	1	9
4	64	16	2	8
9	49	21	3	7
25	16	20	5	4
36	4	12	6	2
16	36	24	4	6
49	1	7	7	1
192	385	147	38	55

$$\bar{X} = \frac{55}{10} = 5.5$$

$$\bar{Y} = \frac{38}{10} = 3.8$$

$$r = \frac{147 - (10)(5.5)(3.8)}{\sqrt{385 - 10(5.5)^2} \sqrt{192 - 10(3.8)^2}} =$$

$$r = \frac{-62}{\sqrt{82.5} \sqrt{47.6}} = -0.989$$

وهذا يعني أن الارتباط قوي وسالب أي أن العلاقة عكسية، فكلما ازداد حجم الأسرة قل مقدار الادخار.

3-3 معامل الارتباط للترتيب والصفات :Coefficient of Rank Correlation

هناك حالات عديدة لا يمكن قياس المتغيرات رقمياً ولهذا توضع على شكل رتب، فمثلاً تحديد لون معين مزج من عدة ألوان فهنا لا يمكن تحديد ذلك اللون بالضبط رقمياً ولكن يمكن أن يرتب حسب اللون وكذلك يمكن تمييز عدة أنواع من الجبن حسب مذاق الملوحة فيه فلا يمكن إعطاء قيم عديدة محددة بالضبط ولهذا اقتربت من أعلى ملوحة إلى أقل على شكل رتب وهذا يشمل متغيرات وظواهر عديدة ومن مختلف الاختصاصات واحد المقاييس الإحصائية المهمة التي تستخدم بشكل واسع لقياس قوة العلاقة هو قانون الرتب والصفات إلى سبيرمان ويرمز له r_s .

Sperman's Correlation Coefficient of Rank

والذي يعرف بالقانون التالي:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

حيث أن n = عدد أزواج المشاهدات.

d_i = تساوي الفرق بين رتب المتغيرات X_i و Y_i .

ويؤدي هذا القانون دوره بشكل معقول ومقبول إذا كان عدد أزواج المشاهدات

أقل من 30 زوجاً.

مثال (3)

أوجد معامل الارتباط للرتب والصفات إلى سبيران لدرجات الطلبة عند التخرج من التوجيهي والتخرج من الجامعة، والتي كانت كما يلي:

X: 56 60 70 85 80 90 82 70 66

Y: مقبول متوسط عالي جيد جداً امتياز جيد متوسط مقبول

درجات X التوجيهي	المستوى Y الجامعي	رتب X_i مع الرتب المعدلة	رتب Y_i مع الرتب المعدلة	d_i	d_i^2
56	مقبول	9	8.5	0.5	0.25
60	متوسط	8	6.5	1.5	2.25
70	متوسط عالي	5.5	5	0.5	0.25
85	جيد	2	3.5	-1.5	2.25
80	جيد جداً	4	2	2	4
90	امتياز	1	1	0	0
82	جيد	3	3.5	-0.5	0.25
70	متوسط	5.5	6.5	-1	1
66	مقبول	7	8.5	-1.5	2.25
					12.5

ولإيجاد معامل الارتباط باستخدام المعادلة أدناه، نحتاج إلى الخطوات التالية:

- 1- ترتيب قيم X_i و Y_i حسب حجمهما.
- 2- استخدام معدل الرتب التي تقابل قيم متساوية بالنسبة لقيم X_i وكذلك Y_i .
- 3- إيجاد الفرق بين رتب X_i ورتب Y_i وإيجاد d_i ثم تربيعها d_i^2 .

$$r_s = 1 - \frac{(6) (12.5)}{(9) (80)} = 1 - \frac{75}{720}$$

$$= 1 - 0.104 = 0.896$$

ولهذا فمعامل الارتباط إلى سبيرمان (r_s) يساوي 0.896.

وهذا يعني وجود علاقة قوية بين درجات التوجيهي ومستوى التخرج من الجامعة.

تطبيقات SPSS:

يمكن إيجاد الارتباط السابق إلى بيرسون وكذلك إلى سبيرمان وغيرها من خلال الإيعازات التالية:

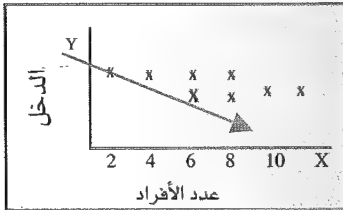
أن نختار Analyze من القائمة الرئيسة للبرنامج SPSS ومن هذه القائمة نحدد Correlate ومن هذه القائمة الفرعية نختار Bivariate ومن هذه الشاشة نحدد الارتباط المطلوب (Spearman و Kendall's و Pearson) بعد تحديد المتغيرات وننقر على ok للتنفيذ.

3-4 الانحدار الخطي البسيط Simple Linear Regression:

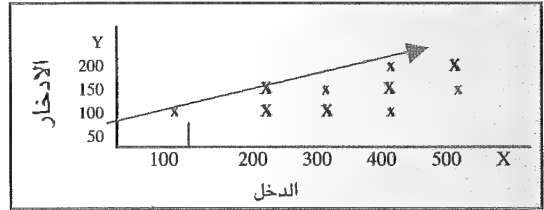
يستخدم الارتباط الخطي البسيط لقياس العلاقة بين متغيرين من النوع المستمر ويمكن أن يكون كلا المتغيرين متغيرات مستقلة أو يكون أحدهما متغير مستقل Independent Variable ويرمز له بالرمز X_i والمتغير الآخر متغير معتمد Dependent Variable ويرمز له بالرمز Y_i .

لقياس معامل الارتباط الخطي البسيط Coefficient of Simple Linear Correlation والذي تم ذكره سابقاً يقيس قوة العلاقة الخطية بين المتغيرين مثل العلاقة بين طول الأب وطول الابن ولكن من الملاحظ هنا أن طول الأب متغير مستقل وطول الابن هو متغير معتمد ونفس الشيء بالنسبة للعلاقة بين درجة التوجيهي ودرجة التخرج من الجامعة أو العلاقة بين الدخل والادخار أو الدخل والمصروف والأمثلة الأخيرة تبين أن هناك علاقة سببية فالمتغير المستقل X_i يتبعه تغير في المتغير المعتمد Y_i وكلما كان التغير متقارب في الكمية وفي نفس الاتجاه يعني ذلك أن العلاقة قوية وموجبة. أما إذا كان التغير في المتغير المستقل X_i يتبعه تغير عكسي في المتغير المعتمد Y_i فيعني ذلك أن العلاقة قوية وسالبة. أما إذا كان التغير في المتغير المستقل X_i لا

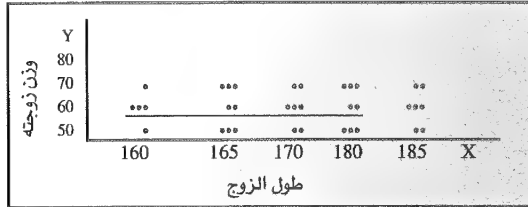
يتبعه تغير في المتغير المعتمد Y_i فهنا لا توجد علاقة بين المتغيرين وهنا يمكن عرض البيانات على شكل الانتشار Scatter plot ومن خلالها نستطيع ملاحظة العلاقة والأشكال الثلاثة التالية تمثل أمثلة لنوع العلاقات بين متغيرات افتراضية بدون بيانات، وهنا الشكل (1) يوضح العلاقة الموجبة بين قيم (X_i, Y_i) لمجموعة من المشاهدات والذي يمثل عوائل تم تسجيل البيانات الخاصة بدخلهم وكذلك ادخارهم الشهري أما الشكل (2) فيوضح العلاقة السالبة بين قيم (X_i, Y_i) للبيانات والذي يمثل مجموعة من العوائل التي تم تسجيل البيانات الخاصة بعدد أفراد العائلة ومقدار الدخل الشهري لذوي الدخل المحدود. أما الشكل (3) فيوضح العلاقة بين متغيرين لا توجد علاقة بينهم. والذي يمثل بيانات طول الزوج ووزن زوجته.



الشكل (2) علاقة سالبة



الشكل (1) علاقة موجبة



الشكل (3) عدم وجود علاقة

وعليه يجب التوضيح بأنه يجب ان يكون هناك تغير منطقي للارتباط فإذا لم يكن هناك تغير منطقي فهذا يعني ان الارتباط حصل بطريقة الصدفة. وهنا يجب التأكيد على ان البيانات يجب ان تتوزع توزيعاً طبيعياً بالنسبة للمتغير المعتمد Y ومعامل الارتباط مقياس مطلق لا يعتمد على الوحدات ولذلك يمكن استخدامه لقياس العلاقة بين متغيرين بوحدات مختلفة.

الانحدار الخطي البسيط يستخدم لوصف العلاقة بين متغيرين أحدهما مستقل X_i والآخر معتمد Y_i من خلال معادلة الانحدار الخطي البسيط والتي هي:

$$Y_i = a + bX_i$$

ومن أهم استخدامات هذه المعادلات هو التنبؤ Prediction بقيم المتغير المعتمد Y_i والتي تسمى بالقيم التقديرية Predicted Values والتي يرمز لها بالرمز y_i من خلال تحديد قيم إلى المتغير المستقل X_i وهذه المعادلة تمثل بياناً بخط مستقيم يمر من خلال الغالبية العظمى لنقاط الانتشار. ومعادلة الانحدار التي تقدر هذه المعادلة هي المعادلة التالية التي تعتمد على العينة وهي: $Y_i = \hat{a} + \hat{b}X_i$ ونعني بالمعادلة الخطية هو أن المتغير المعتمد Y_i هو دالة خطية من المتغير المستقل X_i أي أن زيادة وحدة واحدة إلى X سوف تؤدي إلى التغير في كمية Y_i . وكذلك يتضح أن Y_i هو دالة خطية للثوابت (a, b) والتي تدعى بالمعلمات Parameters ويقال أن النموذج model خطي في معلماته، حيث أن a هو المقطع الصادي constant أي الإحداثي الصادي لنقطة تقاطع خط الانحدار مع Y وأن b هو الميل slope والذي يعني نسبة تغير Y بالنسبة لتغير X وحدة واحدة"، وكمثال آخر على النموذج الخطي في معلماته كما في الصيغة التالية: $(Y = a + b_1X + b_2X_2)$ وهذا النموذج خطي في معلماته (a, b_1, b_2) لكن ليس في المتغير المستقل X والحقيقة أن كلمة خطي تطلق بصورة عامة على النموذج. ويلاحظ بأن بعض النماذج تبدو دوال غير خطية nonlinear functions.

النموذج $Y_i = \exp(a + bX_i)$ هو نموذج غير خطي في كلاهما في المتغير المستقل X وفي المعلمات (a, b) ويمكن تحويل هذا النموذج إلى خطي بأخذ اللوغاريتم الطبيعي إلى جانبي المعادلة ويصبح كما في الصيغة

$\ln Y_i = a + bX_i$ ويمثل نمودجا خطيا. أما إذا لم نستطيع تحويل النموذج إلى خطي فيجب استخدام طرق الانحدار غير الخطي nonlinear regression methods.

وقبل استخدام الانحدار الخطي يجب توفر الشروط التالية في البيانات:

Assumptions of the general linear model

1- المتغير المعتمد Y_i يتوزع توزيعاً طبيعياً Normal Distribution بالوسط u

وتباين σ^2 أي أن $Y_i \sim N(u, \sigma^2)$.

2- الوسط الحسابي للأخطاء العشوائية (e_i) تساوي صفراً والأخطاء هي الفرق

بين القيم الحقيقية للمشاهدات \hat{Y} والقيم التقديرية \hat{Y}_i أي أن $\sum(Y_i - \hat{Y}_i) = 0$ وعدم توفر هذا الشرط يؤدي إلى مشكلة التميز في المقطع الصادي (a) لا تأخذ قيمتها الحقيقية، وهذا يتم تحقيقه بعد إيجاد المعادلة التقديرية $Y_i = a + b X_i$ واستخراج القيم التقديرية \hat{Y}_i .

3- الأخطاء لها تباين ثابت (σ^2) لا يعتمد على قيم المتغيرات المستقلة. ومشكلة عدم تجانس التباين بالنسبة للأخطاء heteroscedasticity تحدث عندما لا يكون تغير الأخطاء منتظم على جهتي خط الانحدار. وهذا دائما يحدث مع البيانات المالية.

4- الأخطاء يجب أن لا تكون مترابطة autocorrelated. وهذا يعني أن الأخطاء يجب أن تكون مستقلة عن موقع المشاهدات في الملف، الخطأ مع إحدى قيم Y يجب أن لا يكون له تأثير على الأخطاء المرافقة إلى قيم Y_i الأخرى. وعند وجود ارتباط بين الأخطاء فإن ذلك يمثل مشكلة العلاقات المتسلسلة للأخطاء. autocorrelated errors وهذا يحدث دائما مع تحليل السلاسل الزمنية.

5- يجب أن يكون التوزيع الاحتمالي إلى الخطأ e_i توزيع طبيعي Normal Distribution بوسط صفر وتباين (σ^2) أي أن $e_i \sim (N(0, \sigma^2))$.

6- القيم بالنسبة للمتغيرات المستقلة تكون ثابتة في تكرار العينات أو يمكن القول بأن قيم المتغيرات المستقلة أما أن تكون مسيطر عليها أو بشكل كامل متوقعة، ويلاحظ عدم توفر هذا الشرط بالنسبة للبيانات في السلاسل الزمنية في بعض الأحيان.

7- في حالات الانحدار المتعدد المتغيرات المستقلة يجب أن لا تكون مرتبطة مع بعضها البعض بشكل كبير لأن ذلك يولد مشكلة الارتباط الذاتي المتعدد multicollinearity وهناك خيار في البرنامج الإحصائي SPSS لمعالجة هذه المشكلة عند تحليل الانحدار وهو الخيار The collinearity diagnostics وهنا يجب السيطرة على الجدول الخاص بالتحليل من خلال تحديد قيم الاحتمال Tolerance حيث يجب أن تكون قيمته أكبر من 0.20 وكذلك عمود عامل

تضخم التباين (VIF) أي Variance-inflation factor وهذا العمود هو المقلوب بالنسبة إلى ما يقابلها من قيم Tolerance في العمود المقابل وهنا قيمة (VIF) يجب أن لا تزيد عن (4) وعندما تكون (VIF=4) فسوف يكون الخطأ المعياري لمعامل الانحدار ضعف وكلما كبرت قيمة VIF تزداد قيمة الخطأ المعياري Standard error إلى معالم الانحدار (b) regression coefficient وهنا يصبح من الصعوبة جدا تقدير أو تقييم قيمة معالم الانحدار (b) وحالات عديدة تغير إشارة معالم الانحدار (b) (Fox, 1991:12) ويعد هذا الشرط من أهم الشروط التي يجب توفرها بالبيانات التي ترغب تحليلها باستخدام أسلوب الانحدار الخطي. وهنا سوف نحسب قيم معالم النموذج الخطي البسيط وكما يلي: $Y = a + b X_i$

حيث أن (b) تمثل الميل وهونسبة تغير قيمة المتغير المعتمد Y إلى تغير وحدة واحدة من المتغير المستقل X.

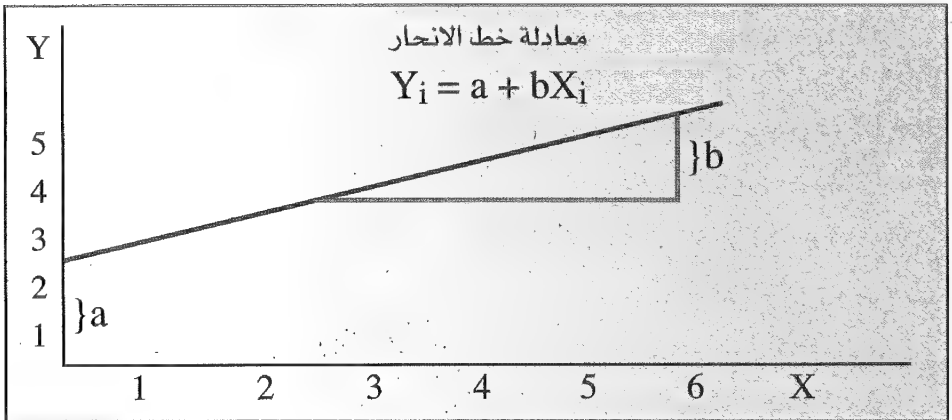
وأن (a) تمثل معامل التقاطع والذي يعني مقدار قيمة Y عندما تكون قيمة المتغير المستقل X تساوي صفرا وفي كثير من الأحيان هناك صعوبة لتفسيرها. أما كيفية حساب كلا من معامل الانحدار باستخدام طريقة المربعات الصغرى Least squares method فهو كما يلي:

$$\hat{b} = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n \bar{X}^2}$$

$$\hat{a} = \bar{Y} - \hat{b} \bar{X}$$

أما تمثيل خط الانحدار وهذه المعاملات في لوحة الانتشار يظهر في الشكل (4)

التالي:



الشكل (4) رسم معادلة خط الانحدار

إذا تم رسم شكل الانتشار ومعادلة الانحدار في مكان واحد. فلدينا ولكل قيمة من القيمة x_i قيمتين مناظرتين هما قيمة حقيقية وهي y_i وتأتي من قيم السلسلة الأصلية وقيمة تقديرية \hat{y}_i تأتي من استخدام معادلة الانحدار المحسوبة عن طريق التعويض عن المتغير x_i بالقيمة المحددة. الفروقات أو ما يسمى بالانحراف error والذي يرمز له بالرمز e_i هي عبارة عن الفرق بين القيمة الحقيقية والقيم النظرية، أي أن:

$$e_i = y_i - \hat{y}_i$$

ومن أهم خصائص خط الانحدار أن $\sum e_i = \sum (y_i - \hat{y}_i) = 0$ وأن

$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2$ هو أقل ما يمكن الحصول عليه من مجموع للمربعات عن استخدام أي معادلة أخرى غير معادلة خط الانحدار المقدرة آنفاً. وبالتالي فإن $\sum e_i^2$ والذي يطلق عليه اسم مجموعات مربعات الخطأ SSE هو المناسب لتقدير ذلك الجزء من جدول تحليل التباين ويستخدم أيضاً لتقدير التباين من استخدام خط الانحدار في التقدير.

مثال (4)

البيانات التالية تمثل الطول والوزن لمجموعة من الأشخاص.

X: الطول	170	172	174	169	170	180	182	183	165	168
Y: الوزن	69	65	76	68	71	78	85	84	64	63

والمطلوب: هل يوجد ارتباط بين المتغير المستقل (X الطول) والمتغير المعتمد (Y الوزن). أوجد المعادلة التقديرية لانحدار Y على X. واختبر هل هناك تأثير معنوي للمتغير المستقل X على المتغير المعتمد Y. ثم أوجد معامل التحديد (R^2) Determinant Coefficient أي نسبة تأثير المتغير المستقل على المتغير المعتمد.

الحل

وكما تم عمله لإيجاد معامل الارتباط r نوجد الأعمدة x_i^2 و Y_i^2 و $x_i y_i$ ومجاميعهم لإيجاد:

$$r = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sqrt{\sum X_i^2 - n \bar{X}^2} \sqrt{\sum Y_i^2 - n \bar{Y}^2}}$$

$$r = \frac{125722 - (10 \times 173.3 \times 72.3)}{\sqrt{300683 - 10(173.3)^2} \sqrt{52857 - 10(72.3)^2}} = 0.937$$

والذي يعني وجود علاقة قوية وموجبة.

أما لإيجاد المعادلة التقديرية باستخدام طريقة المربعات الصغرى فلدينا

$$\hat{Y} = a + b X$$

حيث أن

$$\hat{b} = \frac{SS_{xy}}{SS_{xx}} = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}$$

أو تبسيط الحل كما يلي:

$$\hat{b} = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n \bar{X}^2} = \frac{125722 - (10 \times 173.3 \times 72.3)}{300683 - 10(173.3)^2} =$$

$$= \frac{426.2178}{354.1}$$

$$= 1.203$$

$$a = \bar{y} - b\bar{X} = 72.3 - 1.203 \times 173.3 = -136.238$$

$$\hat{Y} = -136.238 + 1.203X_i$$

حيث أن b^{\wedge} هو معامل الانحدار، وأن أي زيادة وحدة واحدة في الطول (1 سم) تؤدي إلى زيادة الوزن بمقدار (1.203).

وهنا تم استخدام طريقة المربعات الصغرى Ordinary Least Squares Method لإيجاد جدول تحليل التباين (ANOVA) لاختبار هل يوجد تأثير معنوي للمتغير المستقل (X) على المتير المعتمد (Y).

جدول تحليل التباين الأحادي (ANOVA) Table

مصدر التغير S.V. source of variation	مجموع المربعات SS sum of squares	درجات الحرية df Degrees of freedom	متوسط المربعات mss mean squares	قيمة F المحسوبة Computed F
Regression	SSR	1	$MSR = \frac{SSR}{1}$	$= \frac{MSR}{MSE}$
Residual	SSE	(n - 2)	$MSE = \frac{SSE}{n - 2}$	
Total	SST	(n - 1)		

وهنا تم اختبار الفرضية التالية:

فرضية العدم وهي الفرضية المراد اختبارها هي:

$$H_o : b = 0$$

والتي تقول لا يوجد تأثير معنوي للمتغير X على المتغير Y وذلك من خلال مقارنة F المحسوبة مع F الجدولية فإذا كانت قيمة F المحسوبة اصغر من قيمة F الجدولية فهنا نقول أن هذه الفرضية صحيحة.

$$H_1 : b \neq 0 \quad \text{أما الفرضية البديلة فهي:}$$

وهذه الفرضية صحيحة إذا كانت قيمة F المحسوبة اكبر من قيمة F الجدولية.
وجداول تحليل التباين كما يلي (ANOVA)

مصدر التباين s.v.	مجموع المربعات SS	درجات الحرية df	متوسط المربعات ms	F المحسوبة	Sig.
Regression	512.740	1	512.74	57.482	0.000
Residual	71.36	8	8.92		
Total	584.10	9			

هذا يعني هناك تأثير معنوي للمتغير المستقل X_i (الطول) على المتغير المعتمد Y_i (الوزن)، حيث أن قيمة $F = 57.482$ المحسوبة أكبر من F الجدولية وذلك لأن مستوى المعنوية يساوي $(\alpha = 0.000)$ أي أن هناك تأثير معنوي وحقيقي لتغير الطول X_i على الوزن Y ولا يوجد أي خطأ في القرار الذي يقول بأن الطول يحدد الوزن لكل ألف قرار فالخطأ صفر لأن $(\alpha = 0.000)$ وهذا هو رفض فرضية العدم.
وكيفية حساب جدول تحليل التباين (ANOVA) كما يلي:

1. $SSR =$ (مجموع مربعات الانحدار SS Regression)

$$= b \left[\sum X_i Y_i - n \bar{X} \bar{Y} \right]$$

$$= 1.203 \times 426.2178 = 512.74$$

2. $SST =$ (مجموع المربعات الكلي SS Total)

$$= \sum Y_i^2 - n \bar{Y}^2$$

$$= 52857 - 10(72.3)^2$$

$$= 52857 - 52272.9 = 584.10$$

3. $SSE = SS \text{ Total} - SSR$ (مجموع مربعات الخطأ SS Residual)

$$= 584.1 - 512.74 = 71.36$$

أو يمكن استخدام القانون التالي لإيجاد مجموع مربعات الخطأ SSE:

$$SSE = \sum (Y_i - \hat{Y}_i)^2$$

$$SSE = SS_{yy} - b SS_{xy}$$

$$SS_{yy} = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \quad \text{حيث أن:}$$

ولإيجاد معامل التحديد (R^2) Determinant Coefficient والذي يمثل نسبة مجموع مربعات الانحدار (SSR) على مجموع المربعات الكلي (SST) والذي يعني نسبة تأثير المتغير المستقل (X) في المتغير المعتمد (Y) ويمكن حسابه بالمعادلة التالية:

$$R^2 = \frac{SSR}{SST} \times 100\%$$

$$R^2 = \frac{512.74}{584.10} \times 100\% = 87.78$$

وهذا يعني أن 87.78 من الوزن جاء بتأثير مباشر من الطول، أما النسبة الباقية والتي تمثل 12.22 لا يعرف لها مصدر في هذا المثال.

مثال (5)

استخدم بيانات المثال (2) لتقدير معادلة انحدار الادخار على حجم الأسرة مع عمل جدول تحليل تباين ورسم نقاط الانتشار وخط الانحدار.

الحل

بالرجوع إلى المعطيات الواردة في هذا المثال لدينا:

$$n = 10, \bar{X} = 5.5, \bar{Y} = 3.8, \sum x_i = 55, \sum x_i^2 = 385, \sum y_i = 38$$

$$\sum y_i^2 = 192, \sum x_i y_i = 147$$

لذلك فإن:

$$\hat{b} = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

$$= \frac{147 - (55)(38)/10}{385 - (55)^2/10} = \frac{-62}{82.5} = -0.75$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 3.8 - (-0.75)(5.5) = 7.93$$

وبذلك فإن تقدير معادلة الانحدار هي:

$$\hat{y}_i = 7.93 - 0.75x_i$$

وعند التعويض عن x_i بالقيم الواردة في المثال نجد القيم \hat{y}_i والتي تمثل القيم التقديرية وبذلك نحصل على الجدول التالي الذي يضم القيم x_i و y_i و \hat{y}_i وبأخذ الفرق بين y_i و \hat{y}_i نحصل على القيم $e_i = y_i - \hat{y}_i$ ومن ثم بتربيع هذه القيم نحصل على القيم $e_i^2 = (y_i - \hat{y}_i)^2$ كالآتي:

x_i	y_i	\hat{y}_i	e_i	e_i^2
3	6	5.68	0.32	0.1024
5	4	4.18	-0.18	0.0324
10	0	0.43	-0.43	0.1849
9	1	1.18	-0.18	0.0324
8	2	1.93	0.07	0.0049
7	3	2.68	0.32	0.1024
4	5	4.93	0.07	0.0049
2	6	6.43	-0.43	0.1849
6	4	3.43	0.57	0.3249
1	7	7.18	-0.18	0.0324

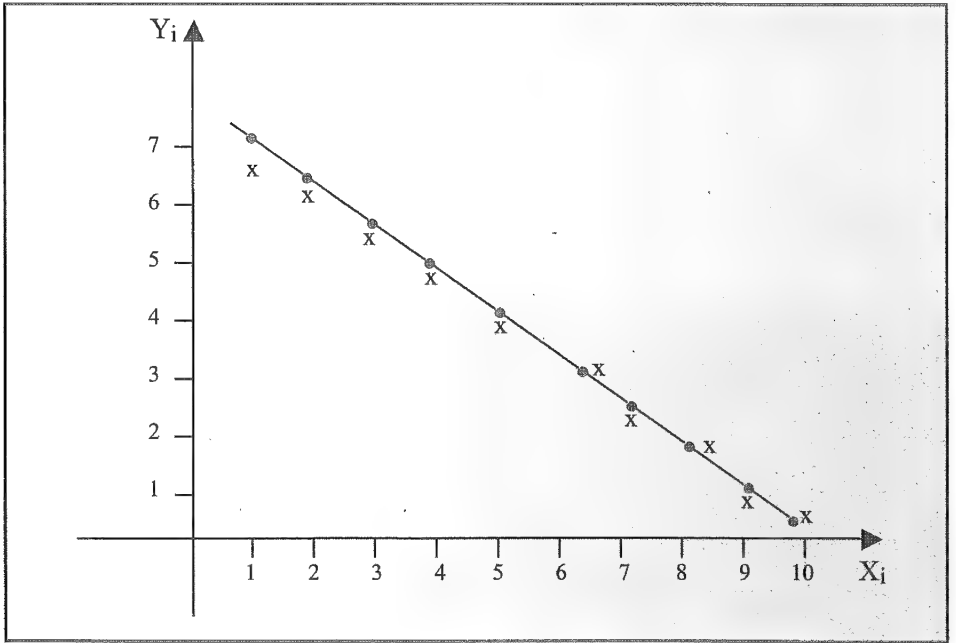
ومن البديهي ملاحظة أن $\sum \hat{e}_i = \sum (y_i - \hat{y}_i) = 0$ مع الأخذ بنظر الاعتبار التقريب للقيم. وكذلك فإن:

$$\sum \hat{e}_i^2 = \sum (y_i - \hat{y}_i)^2 = 1.0065$$

وبذلك فإن تقدير التباين لاستخدام معادلة الانحدار هو:

$$\hat{\sigma}_2 = \frac{\sum \hat{e}_i^2}{n-2} = \frac{1.0065}{8} = 0.1258$$

أما عن رسم نقاط الانتشار فهي رسم القيم x_i مع القيم y_i أما عن تقدير خط الانحدار فهي رسم القيم x_i مع القيم y_i ويظهر الرسمان في الشكل (5) التالي:



الشكل (5) رسم نقاط الانتشار ومعادلة الانحدار

يلاحظ من الشكل (5) أعلاه أن نقاط خط الانحدار تقع على استقامة واحدة وذلك لأنها تمثل رسم معادلة خط مستقيم. فإذا ما تم رسم الخط الواصل بين هذه النقاط نحصل على معادلة الانحدار، والتي إذا ما تم تمديدها على مزيد من البيانات تمثل المعادلة المناسبة لتقدير القيم المستقبلية (التنبؤ).

وبما أن تقدير الادخار لأسرة حجمها (10) أشخاص هو 0.43 دينار فمن البديهي أن الأسر بحجم أكبر لا يستطيعون الادخار.

3-5 العلاقة بين معاملات الانحدار ومعامل الارتباط:

يلاحظ من خلال القوانين السابقة بأن هناك علاقة بين معامل الارتباط r وخط الانحدار $\hat{y} = \hat{a} + \hat{b}x$ من خلال معامل الانحدار \hat{b} والذي يمثل ميل خط الانحدار. أما العلاقة بين r و \hat{b} فتظهر بصورة واضحة من خلال التالي:

كما ذكرنا سابقاً فإن:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

أما

$$\hat{b} = \frac{SS_{xy}}{SS_{xx}}$$

لذلك فإن:

$$r = \sqrt{\frac{SS_{xx}}{SS_{yy}}} \hat{b}$$

أو أن:

$$\hat{b} = \sqrt{\frac{SS_{yy}}{SS_{xx}}} r$$

من خلال هذه العلاقة يمكننا وبسهولة الانتقال بين المفهومين.

نطبيقات SPSS:

يمكن إيجاد خيارات الانحدار Regression كما يلي:

أن نختار Analyze من القائمة الرئيسية ومنها نختار Regression ومن هذا الخيار نختار Linear ونحدد المتغيرات والخيارات المطلوبة وننقر على ok للتنفيذ.

أسئلة

الفصل الثالث

1- إذا توفرت البيانات التالية والتي تمثل X يمثل عدد أفراد الأسرة و Y يمثل عدد قطع الخبز المستهلكة يوميا.

$X:$	4	6	7	3	9	5	8	11	12
$Y:$	10	13	16	8	20	11	17	23	26

المطلوب:

- أوجد معامل الارتباط لبيرسون.
 - أوجد معادلة الانحدار $\hat{Y} = a + bX$
 - قدر عدد القطع المستهلكة (\hat{Y}) إذا كان حجم الأسرة ($X = 17$).
 - أوجد جدول تحليل التباين (ANOVA) واختبر هل أن هناك تأثير معنوي للمتغير المستقل X على المتغير المعتمد Y وبمستوى معنوية 5%.
 - أوجد معامل التحديد.
- 2- البيانات التالية تمثل الدخل الشهري X ومقدار الادخار الشهري لمجموعة من العوائل Y .

$X:$	300	350	400	420	500	600	800	850	900
$Y:$	40	50	70	80	100	120	180	200	220

المطلوب:

- أوجد معامل الارتباط لبيرسون.
- أوجد معادلة الانحدار $\hat{Y} = a + bX$
- قدر عدد قطع المستهلكة (\hat{Y}) إذا كان حجم الأسرة ($X = 1200$)

- أوجد جدول تحليل التباين (ANOVA) واختبر هل أن هناك تأثير معنوي للمتغير المستقل X على المتغير المعتمد Y وبمستوى معنوية 1%.
- أوجد معامل التحديد.

3- البيانات التالية تمثل كميات الأمطار الهاطلة في إحدى المناطق الزراعية (x_i) وكميات الإنتاج الزراعي من محصول القمح (y_i) على فترة 8 سنوات:

X_i (الدم)	5	3	2	4	5	3	1	4
Y_i (الآلاف الأطنان)	95	65	40	70	90	60	25	75

المطلوب:

- إيجاد معامل الارتباط.
- إيجاد معامل ارتباط الرتب.
- تقدير معادلة انحدار y على x .
- تقدير كمية الإنتاج إذا كانت كمية الأمطار الهاطلة هي 7 دسم.
- رسم نقاط الانتشار ومعادلة الانحدار التقديرية.
- تقدير التباين.

4- البيانات التالية تمثل الأضرار الناجمة عن الحرائق (y_i) لمسافة مابين مكان وقوع الحريق وأقرب مركز إطفاء (x_i) كالآتي:

X_i (الأميال)	3.4	1.8	4.6	2.3	3.1	5.5	0.7	3.0	2.6	4.3
Y_i (الآلاف الدولارات)	26.2	17.8	31.3	32.1	275	36.0	14.1	22.3	19.6	31.3

X_i	2.1	1.1	6.1	4.8	3.8
Y_i	24.0	17.3	43.2	36.4	26.1

المطلوب:

- إيجاد معامل الارتباط.
- إيجاد معامل ارتباط الرتب.
- تقدير معادلة انحدار Y على X .
- تقدير الأضرار إذا كان مكان وقوع الحريق يبعد بمقدار 5 أميال.
- رسم نقاط الانتشار ومعادلة الانحدار التقديرية.

الخط

4

الفصل الرابع

الاحتمال والمتغير المشوائي Probability and Random Variable

4-1 مقدمة

4-2 نظرية المجموعات

4-3 التجربة، فضاء العينة والحدث

4-4 الاحتمال ومعناه

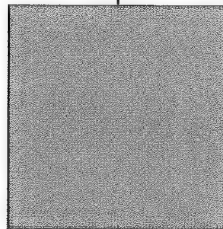
4-5 المتغيرات العشوائية والتوزيعات الاحتمالية

4-6 أمثلة التوزيعات المتقطعة

4-7 أمثلة التوزيعات المستمرة

الفصل الرابع

4



الفصل الرابع الاحتمال والمتغير العشوائي

Probability and Random Variable

4-1 مقدمة Introduction:

يمكن تعريف مفهوم الاحتمال من خلال أعمال نود القيام بها ولا نعرف مدى صحتها او لا نعرف إذا كنا سنستطيع القيام بها أم لا. ولذلك فإن تلك الأعمال تخضع لدرجة من عدم التأكد. وحيث أن الإحصاء الاستدلالي يعتمد على المعلومات الموجودة في العينة لاستخلاص النتائج حول المجتمع، لذلك فإن هذه النتائج لا نستطيع تأكيد صحتها وعدم تأكيد صحة النتائج موروث وموجود في الإحصاء الاستدلالي لذلك قبل الدخول بالأساليب الإحصائية المستخدمة يجب الدخول في دراسة عدم التأكد. العلم الذي يبحث في عدم التأكد هو نظرية الاحتمال Probability Theory والتي تساعدنا في السيطرة على مقدار عدم التأكد وكذلك تساعدنا لتعميم استخدامات المفاهيم التي تصح لمتغير يعتمد على مجتمع محدد ومنها الوسط الحسابي والانحراف المعياري لان تكون مفاهيم تصح لجميع أنواع المتغيرات. وبهذا علينا معرفة مفهوم المتغير العشوائي Random Variable وتوزيعه الاحتمالي Probability Distribution.

لابد قبل الدخول في حساب الاحتمال، من ذكر بعض التعاريف المهمة لمفردات مرافقة لدراسة مفهوم الاحتمال ومنها نظرية المجموعات ومبدأ التجربة والحدث وغيرها سيتم توضيحها كما يلي:

4-2 نظرية المجموعات Set Theory:

المجموعة Set هي تجمع من أشياء Objects أو مفردات Observations تدعى بالعناصر Elements. ولهذه العناصر خصائص مشتركة تجعلها تنتمي لمجموعة او

أخرى و يرمز للمجموعة عادة بالرموز A, B, C, D, \dots او عندما يكون عددها كبيرا نستخدم A_1, A_2, A_3, \dots وحقيقة أن العنصر a يعود او ينتمي للمجموعة A نكتب بالشكل $a \in A$ بينما $b \in A$ يعني أن العنصر b لا ينتمي للمجموعة A .

من أمثلة المجموعات هي مجموعة طلبة الإحصاء الشعبة (1) في كلية الاقتصاد و العلوم الإدارية، مجموعة فريق كرة القدم في الجامعة الأردنية، مجموعة اللجنة الاجتماعية في إحدى النوادي الثقافية وغيرها.

المجموعة يمكن ان تعرض بشكل قائمة من العناصر List of all elements وهذه العناصر تكتب ضمن قوسين $\{ \}$ ، مثلا مجموعة الأعداد الطبيعية (Natural (N Numbers تكتب بالشكل:

$$N = \{1, 2, 3, 4, \dots\}$$

كما يمكن عرض المجموعة بصفة او مؤشر مثل x على أن تحدد قيم ذلك المؤشر، مثلا مجموعة الأعداد الحقيقية الموجبة و لتكن A نكتب بالشكل:

$$A = \{x : 0 \leq x \leq +\infty\}$$

كذلك يمكن عرض المجموعة بشكل مخطط يسمى "مخطط فين" Venn-Diagram حيث تحدد مساحة معينة (دائرة، مستطيل، او أي شكل هندسي آخر) يمثل المجموعة و تكتب العناصر داخله فمثلا المجموعة B و التي تمثل الأرقام الصحيحة الموجبة الواقعة بين الواحد والرقم ستة هي:

B	1	2	3
	•	•	•
	4	5	6
	•	•	•

وبما أننا يمكننا تعريف أكثر من مجموعة فلذلك علينا التعرف على العلاقات التي تحكم وجودها مع بعضها ومن هذه العلاقات نذكر ما يلي:

التساوي Equal: يقال بأن المجموعتين A و B متساويتان إذا كانتا تحتويان على نفس العناصر وعندئذ تكتب $A=B$ مثلا:

$$\{1,2,3\}=\{2,1,3\}=\{3,1,2\}$$

ونعني بذلك أن ترتيب العناصر داخل المجموعة غير مهم . كذلك فإن خاصية التساوي هذه يمكن الاستفادة منها لإيجاد بعض العناصر غير المعلومة في حالة تساوي مجموعتان أو أكثر.

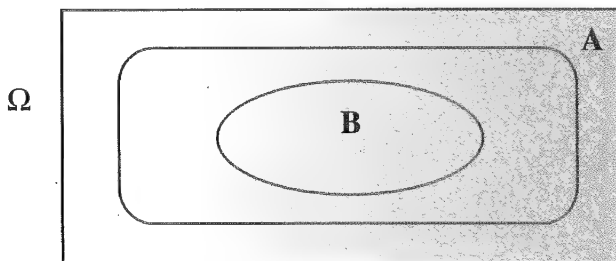
مثلا إذا كانت $A = \{1,2\}$ و $B = \{1,b\}$ فإن العنصر b سيصبح معلوما ويساوي 2، عندما $A=B$.

4

المجموعة الجزئية Subset: المجموعة A هي مجموعة جزئية من المجموعة B وتكتب بالشكل $A \subset B$ إذا كانت جميع عناصر المجموعة A تنتمي أو هي عناصر في المجموعة B بتعبير آخر $A \subset B$ إذا كانت $x \in A \Rightarrow x \in B$ تعطي أن $x \in B$. ويمكن ملاحظة أن:

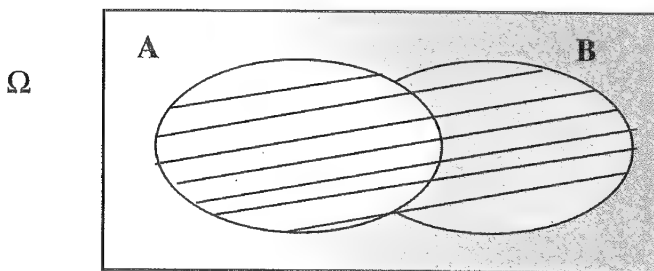
$A=B$ إذا وفقط إذا $A \subset B, B \subset A$ If and only if

يمكن عرض علاقة المجموعات الجزئية بمخطط "فين" كما هو مبين في الشكل (1).



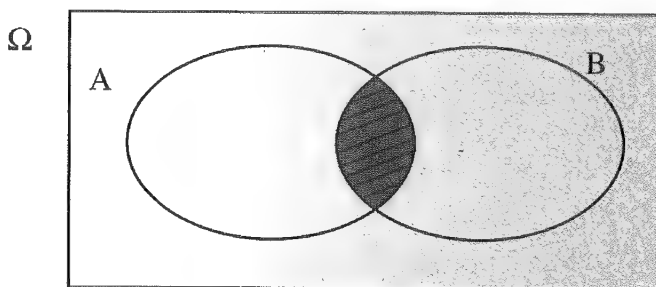
الشكل (1)

الاتحاد Union: اتحاد مجموعتين A و B ويرمز له بالرمز $A \cup B$ يمثل المجموعة التي تحتوي على جميع العناصر التي تنتمي إلى A أو إلى B أو إلى كليهما. ويمكن عرض علاقة اتحاد المجموعتين A و B بمخطط "فين" كما هو مبين بالشكل (2).



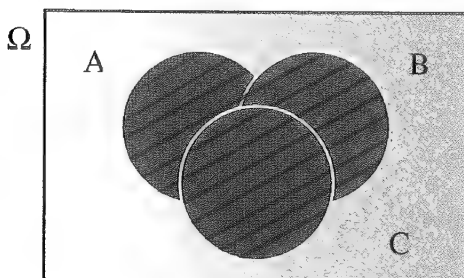
الشكل (2)

التقاطع **Intersection**: تقاطع المجموعتين A و B ويرمز له بالرمز $A \cap B$ يمثل المجموعة التي تحتوي على العناصر المشتركة أي تلك التي تنتمي للمجموعتين A و B معاً، يمكن عرض علاقة تقاطع المجموعتين A و B بمخطط "فين" كما هو مبين في الشكل (3).

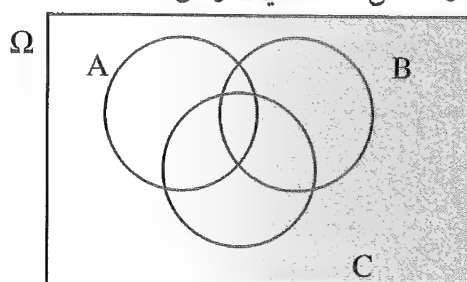


الشكل (3)

وبنفس الأسلوب أعلاه يمكن عرض العلاقات على أكثر من مجموعتين باستخدام مخططات "فين" كما يتضح من الشكلين (4) والذي يعرض $A \cap B \cap C$ والشكل (5) الذي يعرض $A \cup B \cup C$.



الشكل (5)



الشكل (4)

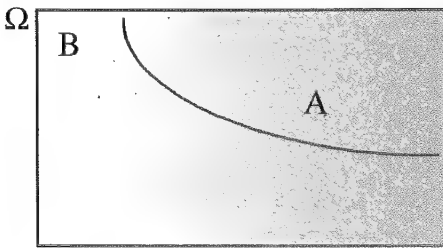
وبما أن المجموعات تظهر مع بعضها بعلاقاتها المختلفة لذلك فيجب أن تكون هناك مجموعة أكبر تحتوي على هذه المجموعات وعلاقاتها.

المجموعة الأكبر تسمى بالمجموعة الشاملة Universal Set والتي تضم جميع العناصر وكذلك جميع المجموعات ويرمز لها بالرمز U or S وفي تطبيقات نظرية المجموعات في الاحتمالات فإن المجموعة الشاملة تناظر ما يسمى بفضاء العينة Sample Space (والذي سيتم التعرف عليها بالفقرة التالية) ويرمز له بالرمز Ω والذي يضم جميع العناصر التي نحصل عليها من التجربة المعينة.

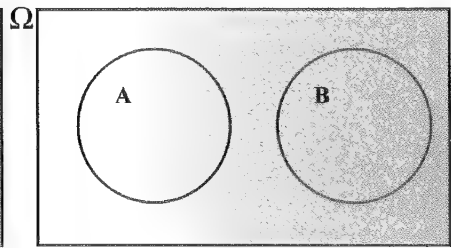
وبالمقابل فإن المجموعة التي لا تحتوي على أي من العناصر والتي تسمى بالمجموعة الخالية Empty Set هي المجموعة الخالية من العناصر ويرمز لها بالرمز \emptyset .

ويتضح من خلال التعاريف أعلاه و وصف المجموعات بأن اصغر مجموعة يمكن التحدث عنها هي \emptyset والتي تكون موجودة او تنتمي لأي مجموعة أخرى مثل A والتي بدورها تنتمي للمجموعة الأكبر Ω وبذلك فإن $\emptyset \subset A \subset \Omega$.

المجموعات المتنافية Mutually Exclusive: يقال بأن المجموعتين A و B متنافيتان إذا وفقط إذا $A \cap B = \emptyset$ ويمكن عرض هذه العلاقة بمخطط "فين" كما هو في الشكلين (6) و (7)



الشكل (7)

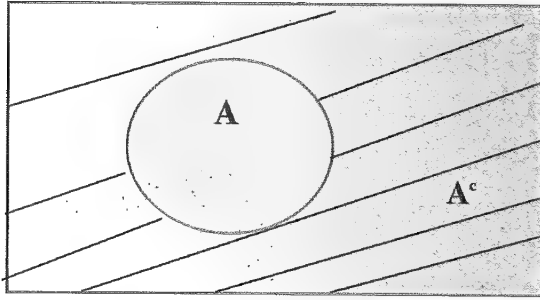


الشكل (6)

ويلاحظ أن الفرق بين الشكلين هو انه في الشكل (6) اتحاد المجموعتين A و B لا يمثل Ω أما في الشكل (7) فإن اتحادهما يمثل Ω .

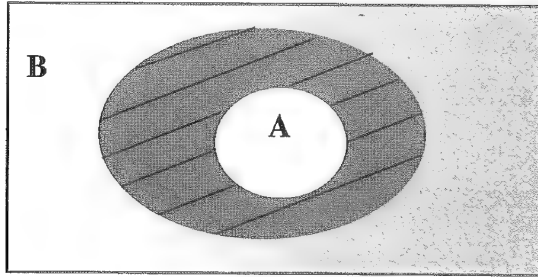
المكملة Complement: لأي مجموعة A تنتمي للمجموعة الشاملة Ω المكملة للمجموعة A ويرمز لها بالرمز \bar{A} أو A^c تعني المجموعة التي تضم العناصر الموجودة

في Ω و غير الموجودة في A . ويمكن عرض هذه العلاقة بمخطط "فين" كما في الشكل (8).



الشكل (8)

الفرق Difference: إذا كانت $A \subset B$ فإن الفرق بين المجموعة الأكبر B والمجموعة الأصغر A والتي يرمز لها بالرمز $B \setminus A$ تمثل العناصر الموجودة في B غير موجودة في A . ويمكن عرض هذه العلاقة بمخطط "فين" كما في الشكل (9)



الشكل (9)

العلاقات على المجموعات:

سيتم في هذه الفقرة عرض بعض العلاقات المهمة على المجموعات والتي يمكن الاستفادة منها ويجب الذكر بأن لهذه العلاقات براهين لن يتم التحدث عنها في الوقت الحاضر إلا أننا سنركز على الجانب التطبيقي للاستفادة من هذه العلاقات في هذا الكتاب.

$$A \cap A = A, \quad A \cup A = A \quad (1)$$

$$A \cap \Omega = A, \quad A \cup \Omega = \Omega \quad (2)$$

$$A \cap \phi = \phi, \quad A \cup \phi = A \quad (3)$$

$$A \cap B = B \cap A, \quad A \cup B = B \cup A \quad (4)$$

وتدعى بقوانين التبادل Commutative Laws

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad (5)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

وتدعى بقوانين التشارك Associative Laws

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad (6)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

وتدعى بقوانين التوزيع Distributive Laws

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \quad (7)$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

وتدعى بقوانين دي مورغان نسبة لاسم العالم De Morgan's Laws

الضرب الكارتيزي Cartesian Products:

الضرب الكارتيزي لـ n من المجموعات A_1, A_2, \dots, A_n هو المجموعة التي تتمثل بالمزدوجات n -tuples معرفة كالاتي:

وأن (x_1, x_2, \dots, x_n) يمثل تجمع مرتب من n من العناصر

Ordered Collection of n Elements

فمثلا لتكن المجموعة $A = \{1, 2, 3\}$ و المجموعة $B = \{4, 5\}$ فإن :

$$A \times B = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}$$

$$B \times A = \{(4, 1), (4, 2), (4, 3), (5, 1), (5, 2), (5, 3)\} \quad \text{أما}$$

$$B \times A \neq A \times B \quad \text{ويمكن ملاحظة أن:}$$

يمكن الاستفادة من خاصية الضرب الكارتيبي للانتقال من مجموعات تمثل نقاط معينة على خط الأعداد الحقيقية أو ما يطلق عليه اسم الإحداثي السيني x -axis إلى مجموعة تمثل نقاط على xy -plane أو ما يطلق عليه اسم الإحداثي السيني x -axis والإحداثي الصادي y -axis. وهكذا

4-3 التجربة، فضاء العينة والحدث:

ولأجل فهم معنى الاحتمال وكيفية حساب الاحتمالات لابد لنا من ذكر بعض المفاهيم الخاصة بهذا الموضوع. وحيث أن نظرية الاحتمالات استحدثت ووطورت لدراسة العلاقات والظواهر العشوائية والتي تخضع للصدف والفرص لذلك سيتم استخدام مفهوم التجربة Experiment وفضاء العينة Sample Space والحدث Event والاحتمال Probability كالآتي:

التجربة Experiment:

والتي تسمى أيضا بالتجربة العشوائية Random Experiment وهي العملية التي نحصل منها على النتائج (out comes) Results أو البيانات Data أو المشاهدات Observations.

ومن أمثلة التجارب العشوائية:

- (1) رمي قطعة نقود وملاحظة الوجه الذي تقع عليه العملة.
- (2) رمي حجر نرد وملاحظة الرقم الظاهر على وجه الحجر.
- (3) اختيار أحد الطلبة وقراءة طوله ووزنه.

فضاء العينة The Sample Space:

وهي المجموعة Set التي تتكون من جميع نتائج أو جميع المشاهدات التي نحصل عليها من التجربة. ويرمز لفضاء العينة عادة بالرمز Ω .

الحدث An Event:

ويمثل مجموعة جزئية Subset من فضاء العينة. ويرمز للحدث عادة بالرموز

A, B, C, \dots او $A_1, A_2, A_3, A_4, \dots$ عندما يكون عددهم كبيرا وقد يكون الحدث:

1 - بسيط Simple event إذا كان مؤلفا من مشاهدة واحدة او نتيجة واحدة

فقط. بتعبير آخر لا يمكن تجزئة الحدث البسيط إلى أكثر من حدث .

2 - مركبا Compound event إذا كان مؤلفا من أكثر من مشاهدة او أكثر من

نتيجة. بتعبير آخر يمكن تجزئة الحدث المركب (الحدث) إلى أكثر من حدث

بسيط.

لأجل استيعاب المفاهيم السابقة فضاء العينة والحدث يمكن الاستعانة

بالأمثلة التالية:

مثال (1)

لرمي قطعة نقود منتظمة اكتب فضاء العينة و الأحداث المكونة لها ؟

الحل

بما أن عملية رمي القطعة المنتظمة و التي لها وجهان هما الصورة Head ونستخدم H والكتابة Tail ونستخدم T لا يمكن أن تكون إلا بحصولنا على إحدى

النتيجتين صورة او كتابة لذلك فإن: $\Omega = \{H, T\}$

أما عن الأحداث المختلفة فلدينا المجموعة او الحدث الخالي \emptyset ثم الحدث البسيط المؤلف من نتيجة واحدة فقط مثل H ثم حدث بسيط آخر مثل T وأخيرا لدينا الحدث المركب من النتيجتين H, T مكونا فضاء العينة Ω لذلك فإن الأحداث الممكنة لهذه التجربة هي :

$$Events : \phi, \{H\}, \{T\}, \Omega = \{H, T\}$$

مثال (2)

لرمي حجر نرد منتظم اكتب فضاء العينة والأحداث المكونة لها ؟

باستخدام نفس الأسلوب الذي اتبع في حل مثال (1) نقول بأن الأرقام 1,2,3,4,5,6 هي جميع نتائج التجربة ولذلك فإنها ستكون فضاء العينة Ω أما الأحداث المختلفة فستبدأ بالأحداث الأصغر \emptyset وتنتهي بالأحداث الأكبر Ω . وتنتج الأحداث من عدد من الأحداث الواقعة فيها بحيث تبدأ بالأحداث التي فيها رقم واحد ثم رقمان ثم ثلاثة وهكذا ، لذلك فإن :

$$\Omega = \{1,2,3,4,5,6\}$$

$$\text{Events} : \emptyset, \{1\}, \{2\}, \dots$$

$$\{1,2\}, \dots$$

$$\{1,2,3\}, \dots$$

$$\{1,2,3,4\}, \dots$$

$$\{1,2,3,4,5\}, \dots, \Omega$$

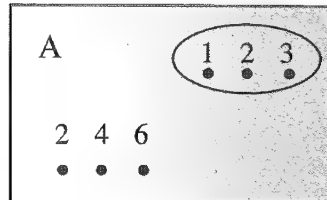
كما يمكن عرض الأحداث بشكل ما يسمى بمخطط "فين" Venn-Diagram والذي يتكون من إطار معين (دائري أو مستطيل) ليُمثل فضاء العينة Ω توضع داخله نقاط لتمثل العناصر المختلفة أو أشكال دائرية لتمثل الأحداث المختلفة. وبذلك فإن مخطط "فين" للمثال الأول هو:

Ω



أما مخطط "فين" للمثال الثاني فهو :

Ω



وإذا افترضنا أن الحدث A يمثل الأرقام الفردية فطريقة عرضه ستمثل الدائرة A في مخطط "فين" أعلاه.

4-4 الاحتمال و معناه Probability & its Meaning

ببساطة الاحتمال هو تعميم لمبدأ النسبة Ratio فعندما يتم اختيار عينة مفردة بطريقة عشوائية من مجتمع محدد (كما رأينا سابقا) فإن الاحتمال هو النسبة . بمعنى أن احتمال اختيار مفردة من مجموعة من n من المفردات هو $1/n$ ولكن التساؤل الذي يطرح نفسه هنا ماذا نعني باستخدام كلمة الاحتمال ونقول بأن الاحتمال القريب من الصفر يعني أننا اقرب ما يمكن بأن الحدث غير محتمل الوقوع بينما الاحتمال القريب من واحد فيعني أننا اقرب ما يمكن بأن الحدث مؤكد الوقوع، ولكن الاحتمال الذي يقع بين اليمتين صفر وواحد فهو الذي نبحث في تفسيره. وبالتالي فعندما نقول بأن الاحتمال مثلا 60%، فإننا نعني بذلك الإمكان النسبي لوقوع ذلك الحدث أي عدد مرات وقوعه بالنسبة لعدد كبير من مرات تكرار تلك التجربة. و يرمز لاحتمال الحدث A بالرمز $P(A)$.

اصبح الآن ممكنا تعريف الاحتمال بالشكل التالي:

الاحتمال Probability

الاحتمال لحدث ما مثل A يعرف على انه الإمكان النسبي Likelihood Relative لوقوع ذلك الحدث و يرمز لاحتمال وقوع الحدث A بالرمز $P(A)$.
خصائص الاحتمال: لأي حدث مثل A ينتمي إلى Ω لدينا :

$$1- P(A) \geq 0$$

$$2- P(\Omega) = 1$$

$$3- A_1, A_2, \dots, A_n \quad \text{ولأي عدد من الأحداث المتنافية}$$

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

ويلاحظ بأن العلاقتين أو الخاصيتين الأولى والثانية واضحة حيث أن التكرار النسبي لا يمكن أن يكون سالبا وبذلك فإن القيم الاحتمالية لا يمكن أن تكون سالبة أما عن الحدث الأكيد Ω فإن احتمال وقوعه فيساوي الواحد الصحيح. أما العلاقة أو الخاصية الثالثة فتسمى بخاصية الجمع Addition Rule.

أما عن طرق تحديد الاحتمالات فهي:

الطريقة الأولى: طريقة الاحتمالات المبدئية A Priori Probability عندما تعين الاحتمالات للأحداث البسيطة المختلفة قبل إجراء التجربة فإن تحديد الاحتمالات يكون مبدئياً.

فمثلاً لتجربة رمي قطعة النقود المنتظمة فإن :

$$\Omega = \{H, T\}$$

$$1/2 \quad 1/2$$

أي أن احتمال الحصول على كتابة T هو $1/2$ وأن احتمال الحصول على H هو $1/2$.

ولتجربة رمي حجر نرد المنتظم لدينا :

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6$$

أي أن احتمال الحصول أي رقم من الأرقام الستة التي تظهر على وجه الحجر هو نفسه ويساوي $1/6$.

وبذلك يلاحظ أن حساب الاحتمال الحدث مثل A يتألف من عدد من الحالات هو:

$$P(A) = \frac{\text{عدد الحالات الممكنة للحدث } A}{\text{عدد الحالات الكلية}}$$

لذلك فعند حساب احتمال الحدث B في تجربة رمي حجر النرد والذي يمثل الحصول على رقم اقل من او يساوي 3 فيتم باستخدام:

$$B = \{1, 2, 3\}$$

$$P(B) = 3/6 = 1/2$$

أما احتمال الحدث C والذي يمثل الحصول على رقم اكبر من 4 فإن:

$$C = \{5, 6\}$$

$$P(C) = 2/6 = 1/3 \quad \text{و أن}$$

أما الطريقة الثانية لحساب الاحتمال فهي طريقة التكرار النسبي.

Relative Frequency (Empirical) Probability

لنفترض أننا وبشكل عشوائي اخترنا أحد العاملين في جامعة عمان الأهلية و
رغبنا بتحديد احتمال أن يكون مشمولاً بالتأمين الصحي. ففي هذه الحالة علينا
الرجوع إلى البيانات المتاحة لأجل الإجابة، ولنفترض أننا وجدنا أن في الجامعة 200
عامل، 120 منهم مشمولين بالتأمين الصحي لذلك فإن:

$$P(A) = \frac{\text{عدد الحالات الممكنة للحدث A}}{\text{عدد الحالات الكلية}} = \frac{f}{n} = \frac{120}{200} = \frac{120}{200} = 60\%$$

ويلاحظ هنا أن $\frac{f}{n}$ والتي تمثل $\frac{\text{عدد الحالات الممكنة للحدث}}{\text{عدد الحالات الكلية}}$ هي
عبارة عن التكرار النسبي Relative frequency للجدول التكرارية السابقة ذكرها.

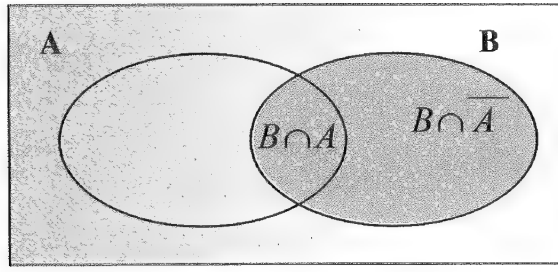
4-4-1 قوانين الاحتمالات:

سيتم هنا ذكر بعض القوانين المهمة بالاحتمالات والتي سيتم الاستفادة منها في
حساب الاحتمالات كما سيتم ذلك تباعاً. هذه العلاقات والقوانين يمكن إيجازها
كآلاتي:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (1)$$

يمكن ملاحظة أن هذا القانون صحيح بالرجوع إلى الشكل (01) مع أن برهان
هذه العلاقة لن يتم ذكره هنا حيث أن:

$$B = (B \cap A) \cup (B \cap \bar{A})$$



الشكل (10)

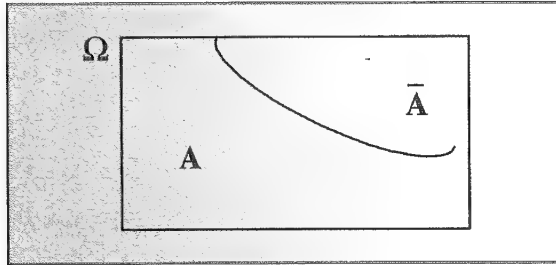
أما إذا كان الحدثان A, B متنافيان فإن $A \cap B = \emptyset$ وبما أن $P(\emptyset) = 0$ فإن :

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \quad (2)$$

$$P(\bar{A}) = 1 - P(A) \quad (3)$$

يمكن ملاحظة صحة العلاقة بالاستعانة بالشكل (11).



شكل (11)

$$A \cup A = \Omega$$

حيث أن

$$P(\Omega) = 1$$

وأن

$$P(\emptyset) = 0 \quad (4) \text{ ونستطيع استنتاج هذه العلاقة من خلال العلاقة :}$$

$$\emptyset \cup \Omega = \Omega$$

$$\emptyset \cap \Omega = \emptyset$$

(5) أما عن تحديد الاحتمال لأي حدث مثل A يتألف من أكثر من حدث بسيط

$$P(A) = \sum_{i=1}^n P(E_i)$$

فإن: $n, \dots, 2, 1 = i, E_i$

فمثلا لحساب احتمال الحصول على رقم فردي عند رمي حجر نرد منتظم فلدينا

الحدث A:

$$A = \{1,3,5\}$$

إذا افترضنا أن $E_1 = \{1\}$, $E_2 = \{3\}$ و أن $E_3 = \{5\}$ فإن :

$$\begin{aligned} P(A) &= \sum P(E_i) \\ &= P(E_1) + P(E_2) + P(E_3) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2} \end{aligned}$$

سيتم الآن استخدام جميع العلاقات المذكورة سابقا لإيجاد احتمال أحداث مختلفة ضمن التطبيقات التالية:

مثال (3)

لتجربة رمي قطعتي نقد منتزمتين. حدد فضاء العينة والاحتمالات للأحداث البسيطة المؤلف منها فضاء العينة.

لنفترض أن الحدث A يمثل الحصول على كتابة واحدة.

والحدث B يمثل الحصول على الأقل على كتابة واحدة.

أوجد $P(B)$, $P(A)$, $P(A \cup B)$, $P(A \cap B)$, $P(B)$, $P(A)$

$$\Omega = \left\{ \begin{array}{cccc} HH & HT & TH & TT \\ 1/4 & 1/4 & 1/4 & 1/4 \end{array} \right\}$$

الحل

وحدث الحصول على كتابة واحدة هو: $A = \{HT, TH\}$

$$P(A) = 1/4 + 1/4 = 1/2 \quad \text{وأن}$$

أما حدث الحصول على الأقل على كتابة فهو $B = \{HT, TH, TT\}$

$$P(B) = 1/4 + 1/4 + 1/4 = 3/4 \quad \text{وأن}$$

$$A \cap B = \{HT, TH\} = A \quad \text{أما عن بعض من الاحتمالات فلدينا ما يلي:}$$

وذلك يعني أن تقاطع الحدثين A, B هو الحدث A وذلك يعطي أن $A \subset B$ مما يعطينا فكرة أن $P(A) < P(B)$ وهو الذي حدث فعلا من ملاحظة الرقمين اللذين تم الحصول عليهما كاحتمال للحدث A وللحدث B .

$$A \cup B = \{HT, TH, TT\} = B \quad \text{أما}$$

وهي علاقة واضحة من خاصية المجموعة الجزئية $A \subset B$ وبذلك نقول بأنه إذا كانت A مجموعة جزئية من B فإن $A \subset B$

$$A \cup B = B \quad \text{وأن} \quad A \cap B = A \quad \text{وذلك يعني أن}$$

$$\bar{A} = \{HH, TT\} \quad \text{وكذلك فإن}$$

$$P(\bar{A}) = 1/4 + 1/4 = 1/2 \quad \text{وأن}$$

$$P(A) + P(\bar{B}) = 1/2 + 1/2 = 1 = P(\Omega) \quad \text{ويلاحظ أن}$$

$$\bar{B} = \{HH\} \quad \text{وأخيرا فإن}$$

$$P(\bar{B}) = 1/4 \quad \text{وأن}$$

$$P(B) + P(\bar{B}) = 3/4 + 1/4 = 1 = P(\Omega) \quad \text{ويلاحظ كذلك بأن}$$

مثال (4)

لتجربة رمي حجري نرد منتظمين. المطلوب:

(أ) حدد فضاء العينة والاحتمالات للأحداث البسيطة المؤلف منها فضاء العينة.

(ب) عرف الحدث A بأنه الحصول على رقم متساو على وجهي حجري النرد أوجد $P(A)$.

(ج) عرف الحدث B بأنه الحصول على رقم 1 على وجه حجر النرد الأول،
أوجد $P(B)$.

(د) عرف الأحداث التالية وأوجد احتمالاتها $A \cup B$, $A \cap B$.

الحل

$$\Omega = \left\{ \begin{array}{l} (1,1), (1,2), \dots, (1,6) \\ (2,1), (2,2), \dots, (2,6) \\ (3,1), (3,2), \dots, (3,6) \\ (4,1), (4,2), \dots, (4,6) \\ (5,1), (5,2), \dots, (5,6) \\ (6,1), (6,2), \dots, (6,6) \end{array} \right\}$$

أي أن هناك 36 حدث بسيط لفضاء العينة، حيث أن كل الأحداث البسيطة لها نفس الاحتمال لذلك فإن:

$$P(E_i) = 1/36, \quad i = 1, 2, \dots, n$$

$$A = \{ (1,1), (2,2), (3,3), (4,4), (5,5), (6,6) \}$$

أما

$$P(A) = 6/36 = 1/6$$

وأن

$$B = \{ (1,1), (1,2), (1,3), (1,4), (1,5), (1,6) \}$$

و

$$P(B) = 6/36 = 1/6$$

وأن

$$A \cap B = \{ (1,1) \}$$

ولحساب التقاطع فإن

$$P(A \cap B) = 1/36$$

وأن

وأخيرا فإن:

$$A \cup B = \{ (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,2), (3,3), (4,4), (5,5), (6,6) \}$$

$$P(A \cup B) = 11/36$$

وأن

4-4-2 طرق العد Counting Techniques:

هناك العديد من الطرق التي بواسطتها نستطيع حصر وعد المفردات التي تعود لحدث ما مثل B وكذلك عد المفردات التي تعود للحدث الأكبر Ω وبذلك نستطيع

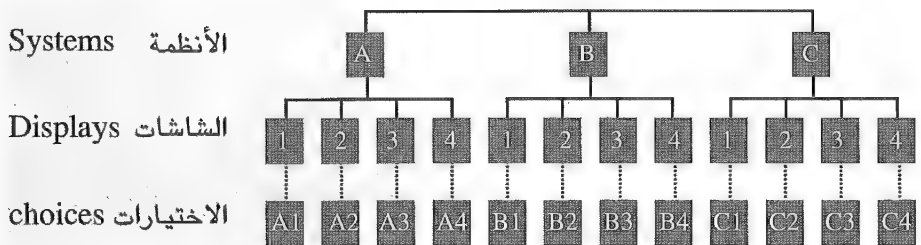
حساب احتمال الحدث B بتقسيم عدد المفردات التي تعود للحدث B على عدد المفردات الكلية. من هذه الطرق نذكر:

(1) طريقة الضرب Multiplication Principle:

وتستخدم هذه الطريقة لحصر وعد المفردات التي تقع في الضرب الكارتيزي لحاصل ضرب مجموعتين او حدثين، وكما تم ذكره سابقا إذا كانت A تحتوي على m من العناصر و B تحتوي على n من العناصر فإن الضرب $A \times B$ له $m \times n$ من العناصر او المفردات المزدوجة. وكتابة هذه المفردات في المجموعة $A \times B$ تحتاج لعمليتين هما اختيار مفردة من المجموعة A لكل ترتيب أول في المفردة المزدوجة واختيار مفردة من المجموعة B لتكون الترتيب الثاني في تلك المفردة المزدوجة. وبما أن هناك m من المفردات في A و n من المفردات في B لذلك فإن m و n تعطي العدد من الطرق لحصر المفردات المزدوجة و بذلك فإن عدد هذه الطرق هو mn.

وبذلك أن طريقة الضرب تذكر كما يلي : إذا كانت العملية الأولى تحتاج إلى m من الطرق ولكل طريقة منها تحتاج العملية الثانية إلى n من الطرق فإن عدد الطرق لإنجاز العمليتين ستحتاج إلى mn من الطرق.

كثير من الكتب تستخدم ما يسمى الشجرة Tree Diagram لتحديد الطرق المختلفة وبافتراض أن شركة لبيع الأجهزة الإلكترونية تستخدم منتجات من ثلاثة أنواع مختلفة من الأنظمة ولتكن A, B, C وتعرض أربعة أنواع مختلفة من شاشات العرض (المونيتور) ولتكن 1, 2, 3, 4 فإن الشخص الذي يرغب بشراء أحد الأجهزة يستطيع أن يختار من بين $3 \times 4 = 12$ من الأجهزة المختلفة. وهذه الطريقة مبينة بالشكل (12) والذي يعرضها بشكل شجرة.



الشكل (12)

(2) الترتيب Ordering :

لنفترض أننا لدينا n من المفردات يراد ترتيبها أو وضعها في n من الأماكن أو المواضع بحيث أن كل مفردة تستعمل ولمرة واحدة فقط، فإن عدد الطرق الممكنة للترتيب هنا يسمى مضروب أو مفكوك n factorial والذي يرمز له بالرمز $n!$ ويمكن تحديد عدد الطرق الممكنة كآلاتي:

n	$n-1$	$n-2$	من المواقع ...	3	2	1
-----	-------	-------	-------------------	---	---	---

لذلك فإن عدد الطرق الممكنة هي :

$$n! = n (n-1) (n-2) \dots 3 \cdot 2 \cdot 1$$

مثال (5)

$$2! = 2 \cdot 1 = 2$$

أوجد:

$$3! = 3 \cdot 2 \cdot 1 = 6$$

$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

وهكذا...

مثال (6)

الأحرف A, B, C يمكن ترتيبها بعدد من الطرق مساوي إلى

$$3! = 3 \cdot 2 \cdot 1 = 6$$

وهذه الترتيبات هي:

ABC, ACB, BAC, BCA, CAB, CBA

مثال (7)

أوجد عدد الترتيبات الممكنة لـ 5 حاسبات إلكترونية. ثم أوجد احتمال أن أحد هذه الترتيبات اختيرت عشوائيا.

عدد الترتيبات او عدد الطرق مساويا إلى

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

أما لحساب الاحتمال فلدينا : $\frac{1}{120}$

(3) التباديل Permutation :

لنفترض أن لدينا n من المفردات يراد ترتيبها أو وضعها في x من الأماكن أو المواضع بحيث أن كل مفردة تستعمل ولمرة واحدة فقط، فإن الطرق الممكنة للترتيب هنا يسمى تباديل x من المفردات من n من المفردات. ويرمز لها بالرمز P_x^n ويمكن تحديد عدد الطرق الممكنة لها كآلاتي:

n	$n-1$	$n-2$	من المواقع ...	$n-x+2$	$n-x+1$
-----	-------	-------	-------------------	---------	---------

لذلك فمن الشكل أعلاه سيبقى لدينا $(n - x)$ من المفردات لن نستخدم لذلك فإن عدد الطرق الممكنة هي :

$$P_x^n = n(n-1)(n-2).....(n-x+2)(n-x+1)$$

من المعلوم أن الضرب او القسمة على حد معين لا يغير المقدار لذلك فإذا تم ضرب الناتج أعلاه في $(n-x)!$ وتمت القسمة عليه نحصل على:

$$\begin{aligned} P_x^n &= \frac{n(n-1)(n-2)....(n-x+2)(n-x+1)(n-x)!}{(n-x)!} \\ &= \frac{n(n-1)(n-2)....(n-x+2)(n-x+1)(n-x)(n-x-1)....3.2.1}{(n-x)!} \\ &= \frac{n!}{(n-x)!} \end{aligned}$$

ويعتبر الأخير صيغة أخرى لحساب قيمة P_x^n وتعتبر اسهل من الصيغة الأولى التي تم ذكرها.

من الأمثلة على استخدام طريقة التبادل لدينا:

مثال (8)

لاختيار الفائز الأول والثاني لعدد من المتسابقين المتقدمين لمسابقة معينة. أوجد عدد الترتيبات الممكنة للاختيار إذا كان عدد المتقدمين 16.

الحل

يلاحظ هنا أن $x=2$ و $n=16$ وبذلك فإن عدد الترتيبات الممكنة هي:

$$P_2^{16} = \frac{16!}{(16-2)!} = \frac{16!}{14!} = \frac{16 \times 15 \times 14!}{14!} = 16 \cdot 15 = 240$$

مثال (9)

لاختيار لجنة من 4 أشخاص من مجموعة مؤلفة من 42 شخص. أوجد عدد الطرق الممكنة لذلك إذا كان لترتيب الأشخاص المختارين أهمية.

الحل

عندما يكون للترتيب أهمية. معنى ذلك نستخدم التباديل لأننا لدينا $x=4$ و $n=42$ وبذلك فإن عدد الطرق الممكنة هي:

$$P_4^{42} = \frac{42!}{(42-4)!} = \frac{42!}{38!} = 42 \cdot 41 \cdot 40 \cdot 39 = 2,686,320$$

مثال (10)

أوجد عدد الطرق الممكنة إلى 6 محاضرين في الرياضيات لتدريس 6 مواد.

$$P_6^6 = 6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$$

$$1) 0! = 1$$

ملاحظة:

$$2) P_n^n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = \frac{n!}{1} = n!$$

(4) التوافيق Combination:

يمثل عدد الطرق الممكنة لتوافيق واختيار x من المفردات من n من المفردات بدون أهمية للترتيب ويرمز له C_x^n أو $\binom{n}{x}$ فيمكن حسابه باستخدام:

$$C_x^n = \frac{n!}{x!(n-x)!}$$

ويمكن فهم مبدأ التوافيق حيث انه يمثل عدد الطرق التي تقسم بها مجموعة من العناصر إلى مجموعتين تحتوي إحداهما عددا معيناً من العناصر وتحتوي الأخرى على بقية العناصر دون النظر إلى ترتيب تلك العناصر في أي من المجموعتين. وكذلك يمكن ملاحظة أن ما يميز التباديل والتوافيق هو أن الترتيب مطلوب وأساسي للمفهوم الأول. ومن الأمثلة على استخدام طريقة التوافيق لدينا:

مثال (11)

8 أشخاص تقدموا للعمل في 4 وظائف مختلفة. فإن عدد الطرق الممكنة

لذلك هو:

$$C_4^8 = \frac{8!}{4!4!} = \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4!}{4 \cdot 3 \cdot 2 \cdot 1 \cdot 4!} = 70$$

مثال (12)

لاختيار حرفين من الأحرف A, B, C, D, and E فإن عدد الطرق الممكنة لذلك

هو:

$$C_2^5 = \frac{5!}{3!2!} = 10$$

مثال (13)

لاختيار لجنة من 4 أشخاص من مجموعة مؤلفة من 10 أشخاص فإن عدد الطرق

الممكنة لذلك هو:

$$C_4^{10} = \frac{10!}{4!6!} = 210$$

ويمكن ملاحظة أن $C_r^n = C_{n-r}^n$ وهي تدعى بمعاملات ذي حدين Binomial

Coefficients

أمثلة أخرى عن مفاهيم الترتيب، التباديل والتوافيق.

مثال (14)

بكم طريقة يمكن ترتيب حروف كلمة Mean ؟

الحل

لكلمة mean أربعة أحرف. و لذلك فإن عدد الطرق هي:

$$4 \quad 3 \quad 2 \quad 1$$

$$4! = 4.3.2.1 = 24$$

مثال (15)

كم كلمة من ثلاثة أحرف يمكن صياغتها من الحروف a, b, c, d, e

الحل

$$P_3^5 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = \frac{5.4.3.2!}{2!} = 5.4.3 = 60$$

مثال
(16)

صف فيه 10 طلاب. بكم طريقة يمكن اختيار لجنة مؤلفة من 3 ؟

الحل

$$C_3^{10} = \frac{10!}{3!7!} = \frac{10.9.8.7!}{3.2.1.7!} = 120$$

مثال
(17)

صندوق فيه 5 كرات بيضاء و 7 كرات حمراء.

(أ) بكم طريقة يمكن اختيار 4 كرات من الصندوق الحل هو:

$$C_4^{12} = 495$$

(ب) بكم طريقة يمكن اختيار 4 كرات كل اثنتين من لون والحل هو:

$$C_2^5 C_2^7 = (10)(21) = 210$$

(ج) بكم طريقة يمكن اختيار 1 بيضاء و 3 حمراء والحل هو:

$$C_1^5 C_3^7 = (5)(35) = 175$$

(د) إيجاد احتمال الفرعين (ب) و (ج) وهما:

210	175
-----	-----
495	495

على التوالي

مثال (18)

لدى مستودع الجامعة 12 حاسبة إلكترونية منها آلتان عاطلتان، تسلمت إحدى الدوائر 4 آلات عشوائيا من هذا المستودع. أوجد:

(1) احتمال عدم وجود آلة عاطلة ضمن ما استلمته الدائرة و الحل هو :

$$\frac{C_0^2 C_4^{10}}{C_4^{12}}$$

(2) احتمال وجود آلة معطلة واحدة ضمن ما استلمته الدائرة و الحل هو :

$$\frac{C_1^2 C_3^{10}}{C_4^{12}}$$

مثال (19)

3 طلاب في غرفة، ما احتمال أن لا يكون من بينهم أي اثنين او اكثر لهم تاريخ الميلاد نفسه؟ أي ما احتمال أن تكون تواريخ ميلادهم مختلفة عن بعضها البعض؟

الحل هو:

$$\frac{(365)(364)(363)}{(365)(365)(365)}$$

سنعرض الآن بعضا من الطرق المفيدة في حساب عدد الطرق الممكنة للترتيب

وهي:

قاعدة (1):

عدد الطرق التي تقسم بها مجموعة فيها n من العناصر إلى k مجموعات جزئية بحيث المجموعة الأولى تحوي n_1 من العناصر والثانية تحتوي على n_2 من العناصر وهكذا فإن:

عدد تباديل مجموعة فيها n من العناصر إذا كان ضمنها n_1 من العناصر المتماثلة و n_2 من العناصر المتماثلة المختلفة عن الأولى، n_k, \dots من العناصر المتماثلة المختلفة عن سابقتها فإن:

$$\frac{n!}{n_1! n_2! \dots n_k!}$$

يمثل عدد الطرق الممكنة

مثال
(20)

عدد الطرق لترتيب كلمة Mississippi هو $\frac{11!}{4!2!4!1!}$ أما لكلمة Layla فهو $\frac{5!}{2!2!1!}$

مثال
(21)

بكم طريقة يمكن ترتيب 8 كتب منها 3 رياضيات و 3 إحصاء و 3 كيمياء، والحل هو: $\frac{8!}{3!3!2!}$

مثال
(22)

بكم طريقة يمكن تقسيم 10 مفردات إلى مجموعتين تضم الأولى 4 مفردات والثانية 6 مفردات. والحل هو:

$$\frac{10!}{4!6!}$$

4-4-3 الاحتمال الشرطي : Conditional Probability

يعرّف الاحتمال الشرطي على أنه احتمال وقوع الحدث A مشروطاً بحدث آخر مثل B قد وقع فعلاً، ويرمز لهذا الاحتمال الشرطي بالرمز $P(A|B)$ ويعرف كالاتي :

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$$

بنفس الأسلوب يمكن تعريف الاحتمال الشرطي للحدث B مشروطاً بأن حدث آخر مثل A قد وقع فعلاً، ويرمز لهذا الاحتمال الشرطي بالرمز $P(B | A)$ ويعرف كآلاتي:

$$P(B | A) = \frac{P(A \cap B)}{P(A)}, P(A) > 0$$

ويمكن ملاحظة من التعريف أعلاه أن:

$$P(A \cap B) = P(A)P(B | A)$$

$$P(A \cap B) = P(B)P(A | B)$$

ويطلق على هذه القوانين ما يسمى بقوانين الضرب Multiplication Rules ويستفاد من هذه الملاحظة الأخيرة أو قانون الضرب في احتساب احتمالات التقاطع إذ علمنا أن الأحداث مشروطة بعضها بالآخر.

ومن خلال التعريف أعلاه يمكن استنباط أهمية الاحتمالات الشرطية وتطبيقاتها وخاصة في حالة أن الحوادث تكون متتابعة الحدوث.

أما تطبيقات الاحتمالات الشرطية وكذلك مفهوم الاستقلالية والذي سيتم شرحه لاحقاً فمتعددة وكثيرة ومتنوعة وسيتم تناول حالات مختلفة في الصفحات التالية.

مثال (23)

بالرجوع إلى بيانات المثال (3) السابق لرمي قطعتي النقد لدينا:

$$P(A) = \frac{1}{2}, P(B) = \frac{3}{4}$$

$$P(A \cap B) = \frac{1}{2}$$

$$P(A) = \frac{1}{2}, P(B) = \frac{3}{4}$$

لذلك فإن:

$$P(A \cap B) = \frac{1}{2}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{1/2}{3/4} = 2/3$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$= \frac{1/2}{1/2} = 1$$

مثال
(24)

بالرجوع إلى بيانات تجربة رمي حجري نرد منتظمين وبافتراض أن:

$$P(A) = \frac{12}{36} \quad P(B) = \frac{6}{36}$$

$$P(A \cap B) = \frac{2}{36}$$

وأن

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/36}{12/36} = \frac{2}{12} = \frac{1}{6}$$

وكذلك

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{2/36}{6/36} = \frac{2}{6} = \frac{1}{3}$$

4-4-4 الأحداث المستقلة Independent Events:

يقال بأن الحدثين A, B مستقلان إذا كان حصول أحدهما لا يؤثر على أو يتأثر

بحصول الآخر، أي أن:

$$P(B|A) = P(B) \quad \text{وأن} \quad P(A) = P(A|B)$$

كذلك يمكن التعبير عن الاستقلالية بين الحدثين A و B كآلاتي: يقال بأن الحدثين A و B مستقلان إذا فقط إذا:

$$P(A \cap B) = P(A) P(B)$$

وسيتيم إيضاح خاصية الاستقلالية واختبارها بالمثال التالي :

مثال (25)

بالرجوع إلى بيانات المثال (3) السابق. هل أن الحدثين A و B مستقلان؟

الحل

لدينا $P(A) = 1/2$ و $P(A|B) = 2/3$ وبما أن $2/3 \neq 1/2$ لذلك فإن A و B غير مستقلان.

كذلك لدينا $P(B) = 3/4$ و $P(B|A) = 1$ وبما أن $3/4 \neq 1$

فإن ذلك يعني أن A و B غير مستقلان .

كما يمكن اختبار الاستقلالية بالاستعانة بفكرة مقارنة احتمال التقاطع مع حاصل ضرب احتمالي الحدثان A و B وبذلك فلدينا :

$$P(A) = 1/2 \quad \text{و} \quad P(B) = 3/4 \quad \text{و} \quad P(A \cap B) = 1/2 \quad \text{وبما أن}$$

$$3/4 \cdot 1/2 \neq 1/2 \quad \text{فإن ذلك يعني أن} \quad A \quad \text{و} \quad B \quad \text{غير مستقلان.}$$

مثال (26)

افترض بيانات الجدول الثنائي التالي الذي يمثل تصنيف عدد من الأشخاص حسب نوعين من التصنيفات: التصنيف الأول فيما إذا كان الشخص مؤمناً عند شركة تأمين (1) او (2) او غير مؤمناً. أما التصنيف الثاني فهو فيما إذا كان الشخص لديه طفل أم لا.

	مؤمن عند شركة تأمين (1)	مؤمن عند شركة تأمين (2)	غير مؤمن	
لديه أطفال	145	85	23	253
ليس لديه أطفال	39	42	52	133
	184	127	75	386

ليكن الحدث A أن يكون الشخص مؤمناً عند شركة (1) فإن $P(A)$ يمكن إيجادها

ليكون:

$$P(A) = \frac{184}{386}$$

أما الحدث B فهو أن يكون للشخص أطفال فإن $P(B)$ يمكن إيجادها ليكون:

$$P(B) = \frac{253}{386}$$

وعند تعريف الحدث C لأن يكون الشخص مؤمناً عند شركة (1) ويكون لديه

أطفال فإن $P(C)$ يمكن إيجادها ليكون:

$$P(C) = P(A \cap B) = \frac{145}{386}$$

لذلك فإن احتمال أن يكون الشخص مؤمناً عند شركة (1) بشرط أن يكون لديه

أطفال سيمثل الاحتمال الشرطي $P(A|B)$ ويمكن إيجادها ليكون:

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{145 / 386}{253 / 386} = \frac{145}{253} \end{aligned}$$

مثال
(27)

صندوق فيه 3 كرات حمراء و9 كرات بيضاء سحبت منه كرة واحدة وتم ملاحظة

لونها ووضعت جانبا ثم سحبت كرة أخرى، أوجد احتمال أن الكرة الأولى بيضاء و الثانية حمراء.

الحل

لنفرض أن A يمثل حدث أن الكرة الأولى البيضاء.

وأن B يمثل حدث أن الكرة الثانية حمراء.

لذلك فإن المطلوب هو حساب $P(A \cap B)$ ، وبما أن السحب الثاني لن يتم إلا بعد عملية السحب الأولى لذلك فإن :

$$P(A \cap B) = P(A) \cdot P(B|A) \\ = \frac{9}{12} \cdot \frac{3}{11}$$

مثال (28)

في مدينة ما اطفائيتان تعملان مستقلتين عن بعضهما البعض، احتمال وصول الأولى إلى مكان الحريق خلال خمس دقائق 0.95 واحتمال وصول الثانية إلى مكان الحريق خلال المدة نفسها يساوي 0.80 ، ما احتمال وصول الاطفائيتين إلى مكان الحريق خلال خمس دقائق.

الحل

لنفترض أن A يمثل حدث ان الأولى تصل إلى مكان الحريق فإن:

$$P(A) = 0.95$$

ولنفترض أن B يمثل حدث أن الثانية تصل لمكان الحريق فإن:

$$P(B) = 0.80$$

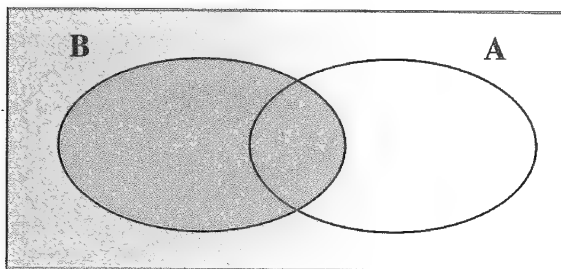
وبما أن الحدثان A و B مستقلان عن بعضهما لذلك فإن احتمال وصول

الاطفائيتين إلى مكان الحريق هو $P(A \cap B)$ ويمثل حاصل ضرب $P(A)$ في $P(B)$ وبذلك فإن:

$$P(A|B) = P(A) P(B) = (0.95)(0.80)$$

نظرية بيز Bayes Theory :

إذا كان Ω يمثل فضاء العينة لتجربة ما، وكان A حدثاً في Ω بحيث $P(A) > 0$ وكان B أي حدث آخر في Ω كما يظهر في الشكل (13).



الشكل (13)

فإن حساب احتمال B يمكن إيجاده بالاستناد إلى خاصية الضرب السابقة و

$$P(B) = P(B \cap A) + P(B \cap \bar{A}) \quad \text{ملاحظة أن :}$$

$$P(B) = P(A)P(B|A) + P(\bar{A})P(B|\bar{A})$$

وبالرجوع إلى قانون الاحتمال الشرطي فإن :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\bar{A})P(B|\bar{A})} \quad \text{وبالتعويض ينتج لدينا :}$$

والأخيرة تسمى قاعدة بيز، في حالة وجود حدثين هما A و B ، ويمكن تعميم النتيجة عند وجود أكثر من حدثين كآلاتي :

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \dots + P(A_n)P(B|A_n)}$$

ولقاعدة او نظرية " بيز " حسب اسم العالم الذي أوجدها " بيز " Bayes أهمية تطبيقية كثيرة ومتعددة وسيتم توضيح إحداها في المثال التالي :

مثال (1)

مصنع فيه 3 ماكنات تصنع كل منها $1/3$ ناتج المصنع. فإذا كان احتمال التالف من إنتاج الماكينة الأولى هو 0.03 ومن إنتاج الثانية 0.02 ومن إنتاج الثالثة 0.04 ، تم عشوائياً سحب عنصر واحد من ناتج المصنع ووجد انه تالفاً أوجد:

1- احتمال انه صنع بواسطة الماكينة الأولى.

2- احتمال انه صنع بواسطة الماكينة الثالثة.

الحل

لنفرض أن: A_1 كون المفردة صنعت بواسطة الماكينة الأولى.

A_2 كون المفردة صنعت بواسطة الماكينة الثانية.

A_3 كون المفردة صنعت بواسطة الماكينة الثالثة.

لذلك فإن $P(A_1) = 1/3$ و $P(A_2) = 1/3$ و $P(A_3) = 1/3$

وبافتراض أن B هو أن تكون المفردة تالفة لذلك فإن هذه المفردة قد تكون مصنعة بواسطة الماكينة الأولى أو الثانية أو الثالثة وذلك يعني أن:

$P(B|A_1) = 0.03$ و $P(B|A_2) = 0.02$ و $P(B|A_3) = 0.04$

وبالتالي فلإيجاد $P(A_1|B)$ باستخدام نظرية "بييز" لدينا:

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)}$$

$$= \frac{(1/3)(0.03)}{(1/3)(0.03) + (1/3)(0.02) + (1/3)(0.04)} = \frac{1}{3}$$

أما لإيجاد $P(A_3|B)$ فإن:

$$P(A_3|B) = \frac{P(A_3)P(B|A_3)}{P(B)}$$

$$= \frac{(1/3)(0.04)}{(1/3)(0.03) + (1/3)(0.02) + (1/3)(0.04)} = \frac{4}{9}$$

4-5 المتغيرات العشوائية و التوزيعات الاحتمالية:

Random Variables and Probability Distributions

كما تم توضيحه سابقا فإن الصفة التي تتغير من حالة إلى أخرى او من مفردة لأخرى توصف ما كان يطلق عليه اسم البيانات سابقا والذي سيوصف ما يطلق عليه هنا اسم المتغير العشوائي. أما عن تعريف المتغير العشوائي فهو دالة Function تمثل العلاقة بين فضاء العينة Ω ومجموعة الأعداد الحقيقية R ولها صفات وخصائص معينة محدودة ولذلك يمكن القول بأن للمتغير العشوائي والذي يرمز له بالرمز X قيمة عددية لكل نتيجة من نتائج التجربة وهناك نوعان من المتغيرات العشوائية:

1- متغير عشوائي متقطع او منفصل Discrete Random Variable:

وهي حالة أن المتغير العشوائي X يمكن كتابة قيمته المتميزة عن بعضها ويظهر بالشكل:

$$X : x_1, x_2, \dots, x_i, \dots$$

وبذلك فمن الممكن للمتغير العشوائي المنفصل أن يأخذ عددا محدودا او معدودا من القيم. من أمثلة هذا المتغير : عدد الطلاب، عدد العلب النالفة في إنتاج معين، عدد حوادث الطرق على أحد التقاطعات، الأرقام التي تظهر على وجه حجر النرد وغيرها.

2- متغير عشوائي مستمر او متصل Continuous Random Variable:

وهي حالة أن المتغير العشوائي X يأخذ مجالا معيناً او حيزاً معيناً على خط الأعداد الحقيقية ويظهر بالشكل التالي $X \in (a, b) \subset R$.

من أمثلة هذا المتغير: الطول، الوزن، كمية الأمطار الساقطة، درجة الحرارة وغيرها.

لكل متغير عشوائي دالة احتمالية مرافقة له يمكن استخدامها لوصف ذلك المتغير وحساب الاحتمالات الخاصة به وإيجاد قيمه المتوقعة وغيرها من المفاهيم، لذلك فعندما نقول متغير عشوائي فإن ذلك يكون ناقصاً من دون قراءة توزيعه الاحتمالي، وسنقوم الآن بدراسة هذا المفهوم كما يلي:

1-5-4 التوزيع الاحتمالي للمتغير العشوائي المنفصل:

Probability Distribution for Discrete Random Variables

التوزيع الاحتمالي للمتغير العشوائي المنفصل يمكن كتابته او توضيحه بشكل جدول Table او رسم Graph او دالة (صيغة) Function (Formula) تمثل او تعطي فكرة كاملة عن قيم ذلك المتغير والاحتمالات المرافقة لتلك القيم. الشكل العام للجدول هو:

$$X: x_1, x_2, \dots, x_i, \dots$$

$$p_1 \quad p_2 \quad \dots \quad p_i \quad \dots$$

مما يعني أن قيم المتغير هي x_1, x_2, \dots على التوالي أما الاحتمالات المرافقة للقيم فهي p_1, p_2, \dots فمثلا القيمة (i) هي x_i أما الاحتمال لها والذي يرمز له بالرمز p_i فهو $P(X = x_i)$ وهكذا، وكذلك يلاحظ ومن خصائص الاحتمالات السابقة أن:

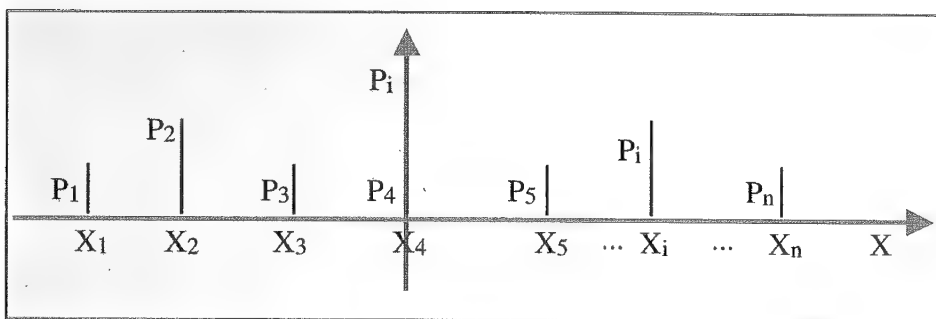
$$\sum_{i=0}^n p_i = 1$$

$$0 \leq p_i \leq 1$$

أما عرض التوزيع الاحتمالي برسم فيتم بوضع نقاط على الإحداثي السيني لتمثل قيم ذلك المتغير المنفصل ورسم أعمدة عند هذه النقاط بأطوال مختلفة تمثل الاحتمالات المناظرة لتلك القيم.

ويمكن ملاحظة أن هذه الطريقة هي طريقة رسم المدرجات التكرارية النسبية السابق عرضها حيث أن الفئات او مراكز الفئات تمثل قيم المتغير المنفصل هنا أما التكرارات النسبية فتمثل الاحتمالات المناظرة.

الشكل العام لرسم التوزيع الاحتمالي يظهر في الشكل (14)



الشكل (14)

الطريقة الأخيرة في عرض التوزيعات الاحتمالية هي صيغة الدالة Function ويمكن القول بأن هذه هي الصيغة الرياضية لعرض التوزيعات الاحتمالية والتي سيتم التعرف عليها من خلال عرض التوزيعات في فقرات لاحقة من هذا الكتاب.

مثال (30)

لتجربة رمي حجر نرد منتظم. عرف المتغير العشوائي X بأنه الرقم الظاهر على الوجه. اكتب التوزيع الاحتمالي للمتغير X .

الحل

يمكن البدء بتحديد Ω بالشكل:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

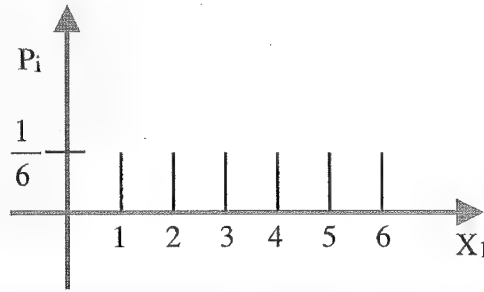
$$1/6, 1/6, 1/6, 1/6, 1/6, 1/6$$

ومن ثم الانتقال إلى صيغة جدول يبين قيم المتغير X واحتمالاته ليصبح:

$$X: 1, 2, 3, 4, 5, 6$$

$$1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6$$

رسم قيم المتغير X واحتمالاته سيظهر بالشكل (15).



الشكل (15) التوزيع الاحتمالي

وأخيرا فإن صيغة الدالة لهذا التوزيع الاحتمالي ستكون :

$$P(X = x_i) = \begin{cases} \frac{1}{6}, & x = 1, 2, 3, 4, 5, 6 \\ 0, & \text{Otherwise (o/w)} \end{cases}$$

يلاحظ بأن الاحتمالات متساوية وثابتة وتساوي $1/6$ ولذلك فإن دالة التوزيع الاحتمالي ستكون الثابت $1/6$ لجميع قيم المتغير والتي هي : 6, 5, 4, 3, 2, 1. أما لحساب بعض الاحتمالات الخاصة بهذا المتغير فمثلا :

$$P(X = 3) = 1/6, P(X = 7) = 0$$

$$P(1 < X \leq 3) = P(X = 2) + P(X = 3) = 2/6$$

$$P(1 \leq X < 3) = P(X = 1) + P(X = 2) = 2/6$$

$$P(1 < X < 3) = P(X = 2) = 1/6$$

$$P(1 \leq X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = 3/6$$

$$P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = 3/6$$

$$P(X < 3) = P(X = 1) + P(X = 2) = 2/6$$

$$P(X \geq 1) = P(X = 1) + P(X = 2) + \dots + P(X = 6) = 1$$

$$P(X > 1) = P(X = 2) + P(X = 3) + \dots + P(X = 6) = 5/6$$

لاحتساب أن يكون المتغير في مجال معين فإننا نحدد النقاط الواقعة ضمن ذلك المجال ونجمع الاحتمالات الخاصة بها.

مثال (1)

لتجربة رمي قطعتي نقد منتزمتين، عرف المتغير العشوائي X بأنه عدد الصور التي تظهر. أوجد التوزيع الاحتمالي للمتغير العشوائي X .

الحل

باتباع الأسلوب السابق لحل المثال نبدأ بتحديد (كالاتي :

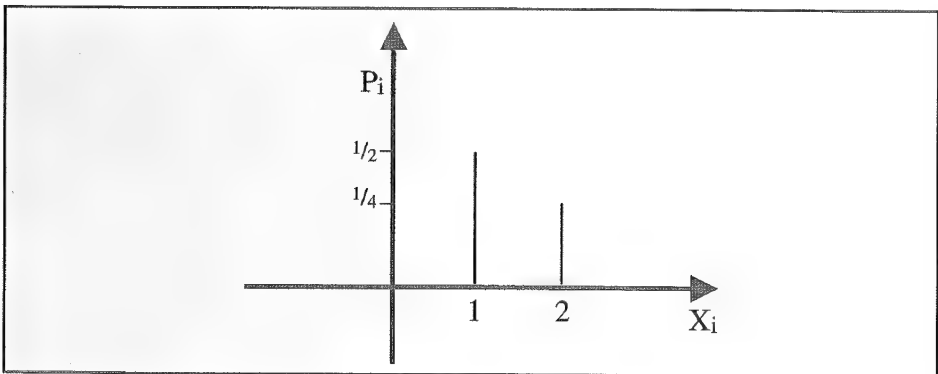
$$\Omega = \{(T,T), (T,H), (H,T), (H,H)\}$$

ولنتقل إلى كتابة التوزيع الاحتمالي بشكل جدول كالاتي :

$$X : 0, 1, 2$$

$$1/4 \quad 1/2 \quad 1/4$$

أما رسم التوزيع فهو في الشكل (16) التالي :



الشكل (16)

وأخيرا فإن صيغة الدالة لهذا التوزيع هي :

$$P(X = x_i) = \begin{cases} \frac{1}{4} & x_i = 0, 2 \\ \frac{1}{2} & x_i = 1 \\ 0 & O / W \end{cases}$$

ففي هذه الحالة الدالة لها قيمتان هما $1/4$ للنقاط 2 و 0 وأن القيمة الثانية هي $1/2$ للنقطة 1 وبذلك نكتب الاحتمال أو التوزيع لعرض هذه الاختلاف كما تم أعلاه.
أما لحساب بعض الاحتمالات فلدينا:

$$P(X=0) = 1/4, \quad P(X=1.5) = 0, \quad P(X=3) = 0$$

$$P(0 < X \leq 1) = 1/4$$

$$P(X \leq 1) = 3/4$$

$$P(0 \leq X < 1) = 1/4$$

$$P(X < 1) = 1/4$$

$$P(0 < X < 1) = 0$$

$$P(X \geq 0) = 1$$

$$P(0 \leq X \leq 1) = 3/4$$

$$P(X > 0) = 3/4$$

يلاحظ من خلال إيجاد الاحتمالات في المثالين السابقين أنه لا احتساب احتمال أن يكون المتغير في مجال معين فإننا نحدد النقاط الواقعة ضمن ذلك المجال ثم نقوم بجمع الاحتمالات الخاصة بتلك النقاط. وهذا يطابق خاصية الجمع التي تم ذكرها سابقاً.

2-4-5-2 التوزيع الاحتمالي للمنفير العشوائي المستمر:

Probability density function for Continuous Random Variable

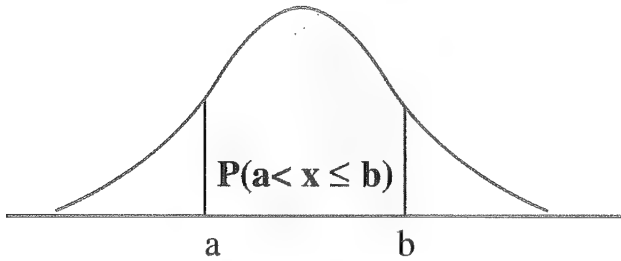
كما تم وصفه سابقاً فإن المتغير العشوائي المستمر أو المتصل continuous random variable يأخذ مجالا معيناً من مجموعة الأعداد الحقيقية \mathbb{R} . ولذلك فإن توزيعه الاحتمالي سيمثل صيغة أو دالة مستمرة تسمى بدالة الكثافة الاحتمالية Probability Density function (p.d.f) تعطي هيئة التوزيع لذلك المجال المعين بحيث أن من خصائص دالة الكثافة والتي يرمز لها بالرمز $f(x)$.

$$\int_{\mathbb{R}} f(x) dx = 1 \quad \text{وأن} \quad f(x) \geq 0 \quad \text{هو أن}$$

ونعني بذلك أن دالة الكثافة تكون دائما موجبة وأن المساحة تحت المنحنى تساوي واحد. أما عن إيجاد الاحتمالات الخاصة بهذا التوزيع فستكون باستخدام الصيغة:

$$P(a < x \leq b) = \int_a^b f(x) dx$$

حيث $-\infty < a \leq b < \infty$ ويمكن القول بأن هذه الصيغ من الاحتمالات تمثل مساحات تحت المنحنى من النقطة a إلى النقطة b كما هو موضح في الشكل (17) التالي:



الشكل (17)

مثال (32)

هل أن الدالة التالية تمثل توزيع احتمالي؟ ولأي نوع من المتغيرات العشوائية؟ ثم أوجد $p(0 < X \leq 1/2)$ حيث أن:

$$f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

الحل

للتأكد من إنها دالة احتمالية يجب توفر شروط الدالة في هذا التطبيق.

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^1 2x dx = x^2 \Big|_0^1 = 1 - 0 = 1 \quad \text{وأن} \quad f(x) \geq 0$$

لذلك فإنها دالة توزيع احتمالي لمتغير مستمر هو X أما الاحتمال فيمكن إيجادها باستخدام صيغة التكامل كآلاتي:

$$p\left(0 < X \leq \frac{1}{2}\right) = \int_0^{1/2} 2x dx = x^2 \Big|_0^{1/2} = \frac{1}{4} - 0 = \frac{1}{4}$$

4-5-3 القيمة المتوقعة (التوقع الرياضي): The Expected value

القيمة المتوقعة للدالة $g(x)$ ويرمز لها بالرمز $E[g(x)]$ يمكن إيجادها باستخدام:

$$E[g(x)] = \begin{cases} \sum_x g(x)p(x) & X \text{ متغير منفصل} \\ \int_{\mathfrak{R}} g(x)f(x)dx & X \text{ متغير مستمر} \end{cases}$$

بشرط أن يكون المجموع أو التكامل موجوداً.

وإذا كان الدالة $g(x) = X$ فإن :

$$E(x) = \begin{cases} \sum_x xp(x) & X \text{ متغير منفصل} \\ \int_{\mathfrak{R}} xf(x)dx & X \text{ متغير مستمر} \end{cases}$$

والأخير يمثل القيمة المتوقعة للمتغير العشوائي X . وتقابل هذه القيمة الوسط الحسابي الذي تم التعرف عليه سابقاً والذي يقابل جمع حواصل ضرب مراكز الفئات بالتكرارات النسبية ليصبح جمع حواصل ضرب قيم المتغير في احتمالاته المقابلة في حالة أن المتغير العشوائي X هو متغيراً منفصلاً.

لإيجاد القيمة المتوقعة للمتغير X في حالتيه المنفصل والمستمر سترجع للمثالين الذين تم ذكرهما سابقاً كآلاتي :

مثال (33)

$$X = -1, 0, 1$$

$$1/4 \quad 1/2 \quad 1/4$$

$$E(x) = \sum x_i p(x_i)$$

فإن

$$= (-1)(1/4) + (0)(1/2) + (1)(1/4) = 0$$

مثال (34)

$$f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

للدالة

$$\begin{aligned} E(X) &= \int xf(x)dx \\ &= \int_0^1 x \cdot 2x dx = 2 \int_0^1 x^2 dx \\ &= \left. \frac{2}{3} x^3 \right|_0^1 = \frac{2}{3} \end{aligned}$$

فإن

ويلاحظ أن اختيار $g(x)$ لتكون أي دالة من المتغير X مناسب لإيجاد الوسط كما رأينا أعلاه. وكذلك لإيجاد التباين فيما لو تم اختيار $g(x)$ لتكون $(x - \mu)^2$ وبذلك فإن حساب التباين σ^2 للمتغير X والذي يرمز له بالرمز $\text{var}(X)$ سيكون باستخدام:

$$\text{var}(X) = \sigma^2 = \begin{cases} \sum (x - \mu)^2 p(x) & \text{متغير منفصل} \\ \int_{\mathfrak{R}} (x - \mu)^2 f(x) dx & \text{متغير مستمر} \end{cases}$$

كذلك يمكن الاستعانة ببعض العلاقات والنظريات التي تتحكم في إيجاد وسط وتباين دوال معينة من المتغير او المتغيرات العشوائية. وسنم ذكر بعضها والذي له أهمية تطبيقية في هذا الكتاب وبدون ذكر البراهين الخاصة بها أما البعض الآخر

والذي يتطلب خلفية رياضية أكثر مما لدى قارئ هذا الكتاب فلن يتم التطرق أو الحديث عنها. لذلك فلدينا هنا:

نظرية Theorem :

إذا كان b و a أعداد حقيقية ثابتة وان

$$\mu_y = E(X) = a + b\mu_x = a + bE(X) \quad \text{فإن}$$

$$\sigma_y^2 = Var(Y) = b^2 Var(X) = b^2 \sigma_x^2 \quad \text{وأن}$$

$$\sigma_y = \sqrt{b^2 \sigma_x^2} = |b| \sigma_x \quad \text{وأخيرا فإن}$$

إذا كانت هذه التوقعات موجودة.

مثال (35)

إذا كان المتغير X له وسط حسابي $\mu_x = 24$ وانحراف معياري $\sigma_x = 5$ وأن المتغير y هو $y = -15 + 4x$ فإن وسط المتغير y وهو μ_y سيصبح:

$$\mu_y = -15 + 4 \mu_x = -15 + 4 (24) = 81$$

أما الانحراف المعياري للمتغير y فهو μ_y سيصبح:

$$\sigma_y = |4| \sigma_x = 4 (5) = 20$$

مثال (36)

إذا كان المتغير X له وسط μ وانحراف معياري σ . اثبت أن المتغير

$$Z = \frac{x - \mu}{\sigma} \quad \text{له وسط صفر وانحراف معياري واحد.}$$

$$Z = \frac{x - \mu}{\sigma} = \frac{x}{\sigma} - \frac{\mu}{\sigma}$$

الحل

وبافتراض أن $a = 1/\alpha$ و $b = -\mu/\alpha$ فإن المتغير Z سيمثل دالة خطية من المتغير x بالشكل $Z = b + aX$ ولذلك فإن:

$$E(Z) = \frac{1}{\sigma} E(X) - \frac{\mu}{\sigma} = \left(\frac{1}{\sigma}\right)\mu - \frac{\mu}{\sigma} = 0$$

$$\text{var}(Z) = \frac{1}{\sigma^2} \text{var}(x) = \frac{\sigma^2}{\sigma^2} = 1 \quad \text{وأن}$$

نظرية:

إذا كان المتغير X له قيمة متوقعة هي μ_x وانحراف معياري هو σ_x وبافتراض أن $k > 1$ تمثل أي عدد ثابت فإن:

$$p(|x - \mu_x| < k\sigma_x) \geq 1 - \frac{1}{k^2}$$

وهذه النظرية يطلق عليها اسم متباينة تشبشف Chebyshev Inequality وتساعدنا في احتساب احتمالات المجال الذي يبعد بـ k من الانحرافات المعيارية عن الوسط وبغض النظر عن التوزيع الاحتمالي للمتغير X .

ويمكن هنا الرجوع إلى تعريف الانحراف المعياري كمقياس لتشتت عينة والقول بأنه ولأي توزيع للمتغير X يمكن حساب احتمالات الفترات التي بواسطتها نبتعد عن قيمة الوسط μ_x .

مثال (37)

إذا كان المتغير X له وسط 24 وانحراف معياري 5 أوجد نسبة القيم التي تقع ضمن الفترة التي تمثل 2 من الانحرافات المعيارية حول الوسط؟

الحل

باستخدام متباينة تشبشف لدينا :

$$P(|x - \mu_x| < 2\sigma_x) \geq 1 - 1/2^2 = 1 - 1/4 = 3/4$$

و تعني على الأقل 75% من القيم ستقع على بعد 2 انحراف معياري حول الوسط.

نظرية

إذا كانت a_1 و a_2 و $a_n \dots$ ثوابت و X_1 و X_2 و $X_n \dots$ متغيرات عشوائية بتوزيع احتمالي معين، وافترض أن المتغير Y هو:

$$Y = \sum a_i X_i \quad \text{فإن}$$

$$E(Y) = \sum a_i E(X_i)$$

$$\text{var}(Y) = \sum_{i=1}^n a_i^2 \text{var}(X_i) + 2 \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{cov}(X_i, X_j)$$

حيث أن $\text{cov}(X_i, X_j)$ يمثل التباين المشترك Covariance بين المتغيرين X_i, X_j . وهذه النظرية لها أهميات تطبيقية كثيرة وخاصة عند عملية التحويل Transformation من عدد من المتغيرات إلى متغيرات أخرى.

وإذا افترضنا أن $n = 2$ لذلك فلدينا $Y = a_1 X_1 + a_2 X_2$ وبافتراض أن $a_1 = 1, a_2 = 1$ يصبح لدينا الشكل المبسط من هذه الدالة وهو:

$$Y = X_1 + X_2$$

$$E(Y) = E(X_1) + E(X_2) \quad \text{لذلك فإن}$$

$$\text{var}(Y) = \text{var}(X_1) + \text{var}(X_2) + 2 \text{cov}(X_1, X_2)$$

نظرية

إذا كانت a_1, a_2, \dots, a_n ثوابت و X_1, X_2, \dots, X_n متغيرات عشوائية مستقلة عن بعضها البعض وافترض أن Y هو:

$$y = \sum_{i=1}^n a_i X_i$$

فإن:

$$E(y) = \sum_{i=1}^n a_i E(X_i)$$

وأن:

$$\text{var}(y) = \sum_{i=1}^n a_i^2 \text{var}(X_i)$$

ما نعينه بالمتغيرات المستقلة Independent variables أن التباينات المشتركة تكون معدومة أي أنه إذا كان المتغيران X و Y مستقلان فإن: $\text{cov}(X, Y) = 0$

مثال (38)

أوجد وسط وتباين المتغير $y = 2X_1 + X_2 - 5$ إذا علمت أن X_1 له وسط 4 وتباين 9 أما X_2 له وسط -2 وتباين 5. علما أن X_1 و X_2 مستقلين عن بعضهما.

الحل

$$Y = 2X_1 + X_2 - 5$$

$$\begin{aligned} E(y) &= E(2X_1 + X_2 - 5) \\ &= 2E(X_1) + E(X_2) - 5 \\ &= 2(4) + (-2) - 5 \\ &= 1 \end{aligned}$$

$$\text{var}(Y) = \text{var}(2X_1 + X_2 - 5)$$

أما

$$\begin{aligned} &= 4 \text{var}(X_1) + \text{var}(X_2) \\ &= 4(9) + 5 = 41 \end{aligned}$$

مثال (39)

إذا علمت أن X_1, X_2, \dots, X_n تمثل n من المتغيرات العشوائية المستقلة والمتشابهة التوزيع بوسط μ وتباين σ^2 . أوجد وسط وتباين \bar{X} .

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$= \frac{1}{n} (X_1 + X_2 + \dots + X_n)$$

$$E(\bar{X}) = \frac{1}{n} [E(X_1) + E(X_2) + \dots + E(X_n)]$$

إذن

$$= \frac{1}{n} [\mu + \mu + \dots + \mu] = \frac{n\mu}{n} = \mu$$

$$Var(\bar{X}) = \frac{1}{n^2} [Var(X_1) + Var(X_2) + \dots + Var(X_n)]$$

أما

$$= \frac{1}{n^2} [\sigma^2 + \sigma^2 + \dots + \sigma^2] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

4-6 أمثلة التوزيعات المتقطعة:

لكون أن المتغيرات العشوائية لها تطبيقات متعددة مختلفة اتفق على استخدام توزيعات احتمالية متعددة تخدم هذه التطبيقات، ومن هذه التطبيقات تجارب ذي الحدين وبرنولي، توزيع بواسون، التوزيع المنتظم، التوزيع الهندسي والفوق الهندسي وذي الحدين السالب، وسيتم إعطاء فكرة عن كل من هذه التوزيعات كآلاتي:

4-6-1 تجربة ذي الحدين Binomial Experiment:

ذكرنا سابقاً أن التجارب العشوائية هي عمليات نحصل منها على النتائج وتجربة ذي الحدين هي نوع من هذه التجارب بالخصائص التالية:

(1) التجربة تتألف من عدد من المحاولات ونفرضه n والذي يمثل حجم العينة.

(2) أن المحاولات مستقلة عن بعضها.

(3) لكل محاولة نتيجتين فقط هما النجاح والfailure.

4) احتمال النجاح متساوي وثابت لجميع المحاولات. وسيتم افتراض أن احتمال نجاح المحاولة هو p وبذلك فإن احتمال فشلها سيكون $1-p$ والذي يرمز له بالرمز q .

وبافتراض أن X هو المتغير العشوائي الذي يمثل عدد حالات النجاح لمثل تلك التجارب فإننا نحصل على قيم لهذا المتغير العشوائي واحتمالات مناسبة له بالشكل الذي يشير إلى أن هذا المتغير هو من النوع المنفصل (المتقطع).

عندما يكون عدد المحاولات n مساوياً للواحد فإن التجربة تسمى تجربة برنولي Bernoulli (حسب العالم الذي أوجدها)، أما عندما يكون n أكبر من واحد فإن التجربة تدعى بتجربة ذي الحدين Binomial experiment او متغير ذي الحدين Binomial Random Variable.

ويرمز لهذا التوزيع بالشكل:

$$X \sim B_i(n, p)$$

حيث أن $n \geq 1$ ويمثل حجم العينة او عدد المحاولات.

وأن $0 \leq p \leq 1$ يمثل احتمال نجاح المحاولة.

أما التوزيع الاحتمالي لهذا المتغير فهو:

$$P(X = x) = \begin{cases} C_x^n p^x (1-p)^{n-x}, & x = 0, 1, 2, \dots, n \\ 0, & \text{otherwise} \end{cases}$$

وأن الوسط والتباين لهذا المتغير هما على التوالي:

$$E(X) = np$$

$$\text{ar}(X) = np(1-p)$$

ويمكن ملاحظة أهمية هذا التوزيع في التطبيقات من خلال الأمثلة التالية:

مثال (40)

افترض تجربة رمي قطعة نقود منتظمة n من المرات ولتكن $n = 2$ ، وافترض أن المتغير العشوائي X يمثل عدد مرات الحصول على صورة. اكتب التوزيع الاحتمالي.

الحل

لحل هذا المثال علينا الرجوع إلى فضاء العينة ومن ثم الانتقال إلى عمل جدول لتوزيع المتغير x كما تم عمله في مثال سابق يعرض هذه الفكرة، وهي طريقة مناسبة وعملية خاصة عندما يكون حجم العينة n صغيراً، أما عندما يكون كبيراً فسيصعب كتابة فضاء العينة وقيم المتغير ومن ثم الاحتمالات ولذلك فإن اتباع صيغة التوزيع الاحتمالي ستكون أفضل وبالشكل التالي:

بما أن $n = 2$ ، $p = 1/2$ حيث أن القطعة منتظمة. لذلك فإن:

$$X \sim B_i(2, 1/2)$$

وعند التعويض في دالة التوزيع الاحتمالي للمتغير ذي الحدين نحصل على

$$P(X = x) = \begin{cases} C_x^2 \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{2-x}, & x = 0, 1, 2 \\ 0, & \text{otherwise} \end{cases}$$

ولذلك فإن إيجاد قيم و احتمالات المتغير ستكون:

$$P(X = 0) = C_0^2 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^2 = 1/4 \quad \text{عندما } x = 0 \quad \text{فإن}$$

$$P(X = 1) = C_1^2 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) = 1/2 \quad \text{عندما } x = 1 \quad \text{فإن}$$

$$P(X = 2) = C_2^2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^0 = 1/4 \quad \text{عندما } x = 2 \quad \text{فإن}$$

مثال (41)

لتجربة رمي خمس قطع نقود منتظمة، اكتب التوزيع الاحتمالي للمتغير X الذي يمثل عدد الصور.

الحل بالرجوع للوصف الذي تم عرضه للمثال السابق فإن:

$$X \sim B_i(5, 1/2)$$

حيث أن

$$P = \frac{1}{2}, \quad n = 5$$

$$P(X = x) = \begin{cases} C_x^5 \cdot \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x}, & x = 0, 1, 2, 3, 4, 5 \\ 0, & \text{otherwise} \end{cases} \quad \text{وأن}$$

وبذلك فإن إيجاد أي قيمة احتمالية سيكون بعملية التعويض في الدالة الاحتمالية ومن هذه الاحتمالات وباستخدام الصيغ الاحتمالية السابقة فإن:

$$P(X = 3) = C_3^5 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = \frac{10}{32}$$

$$P(X = \frac{3}{2}) = 0$$

$$P(2 < X \leq 4) = P(X = 3) + P(X = 4)$$

$$= C_3^5 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 + C_4^5 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^1 = \frac{10}{32} + \frac{5}{32} = \frac{15}{32}$$

مثال
(42)

إذا علمت أن نسبة القطع التالفة في إنتاج معين هو 10% . فإذا تم مراقبة الخمس قطع القادمة أوجد احتمال إيجاد ثلاثة قطع تالفة.

الحل

نفترض أولاً أن المتغير العشوائي X يمثل عدد القطع التالفة، وبما أن $n = 5$ حيث

تم مراقبة خمس قطع إنتاجية وأن $p = 0.10$ وذلك لأن نسبة القطع التالفة أو احتمال أن تكون القطعة تالفة هو 10% لذلك فإن:

$$X \sim \text{Bi}(5, 0.10)$$

وأن التوزيع الاحتمالي هو:

$$P(X = x) = \begin{cases} C_0^5 (0.10)^x (0.90)^{5-x}, & x = 0, 1, 2, 3, 4, 5, \\ 0 & , \text{ otherwise} \end{cases}$$

لذلك فإن احتمال إيجاد ثلاثة قطع تالفة هو عندما $x = 3$ ، أي أن:

$$P(X = 3) = C_3^5 (0.10)^3 (0.90)^2 = 0.0081$$

4-6-2 توزيع بواسون Poisson Distribution:

وهو التوزيع المناسب للمحاولات التي لا يكون لها حد أعلى وكذلك للحالات التي تظهر فيها الحاجة لتحديد عد المرات أو الحالات لفترة زمنية مثلاً عدد المكالمات المستلمة في بدالة معينة لفترة ساعة مثلاً أو عدد الزبائن اللذين تم تقديم الخدمة لهم لفترة عشر دقائق مثلاً، توزيع بواسون بالوسط λ له دالة احتمالية بالشكل:

$$P(X = x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, \dots \\ 0 & , \text{ otherwise} \end{cases}$$

عندما تكون n كبيرة و p احتمال النجاح صغيرة لتوزيعات ذي الحدين فإن التقريب المناسب هو استخدام وسط توزيع بواسون بالحد λ حيث أن $\lambda = np$. كما ويلاحظ وجود جداول خاصة لحساب الاحتمالات الخاصة بهذا التوزيع.

أما عن وسط و تباين توزيع بواسون فهما:

$$\sigma^2 = \lambda, \mu = \lambda$$

إذا كان أحد البنوك يستلم بمعدل 6 شيكات بدون رصيد في اليوم. أوجد احتمال أنه سيستلم 4 شيكات بدون رصيد في يوم معين.

الحل

إذا افترضنا أن $x = 4$ وأن $\lambda = 6$ فإن استخدام دالة التوزيع تعطينا حساب الاحتمال حيث أن:

$$P(X = 4) = \frac{6^4 \cdot e^{-6}}{4!} = 0.134$$

3-4-6 التوزيع فوق الهندسي Hypergeometric Distribution

لو افترضنا أننا نود إيجاد عدد الوحدات المعطوبة في عينة عشوائية مؤلفة من n من الوحدات المسحوبة من مجموعة من N من الوحدات المنتجة من بينها a من الوحدات المعطوبة فإن العينة سيتم سحبها بطريقة حيث أن أي من محاولات السحب المتتالية الوحدات المتبقية سيكون لها نفس احتمال السحب.

وبذلك فإن احتمال أن القطعة الأولى المسحوبة ستكون معطوبة هو a/N ، أما في السحبة الثانية فهو $\frac{a-1}{N-1}$ أو $\frac{a}{N-1}$ معتمدا على ما تم الحصول عليه في السحبة الأولى.

لذلك فإن المحاولات في هذه الحالة ستكون غير مستقلة عن بعضها، كما كان الحال بالنسبة لتوزيع ذي الحدين علاوة على أن توزيع ذي الحدين يستخدم في حالة السحب بالإرجاع Sampling with replacement أما لحل حالات السحب بدون إرجاع Sampling without replacement فسيكون كما يلي:

عدد حالات النجاح (القطع المعطوبة) x سيتم اختيارها من الوحدات المعطوبة a من المجتمع بعدد من الطرق مساويا إلى $\binom{a}{x}$ أما الفشل وهو $x - n$

ويمثل القطع غير المعطوبة فسيتم اختيارها من الوحدات غير المعطوبة $N - a$ من المجتمع بعدد من الطرق مساويا إلى $\binom{N-a}{n-x}$

ولذلك فإن كل الطرق الممكنة للسحب سواء كانت الوحدات معطوبة أم لا فسيتم بعدد من الطرق مساويا إلى $\binom{a}{x} \binom{N-a}{n-x}$

وسينسب إلى عدد الطرق الكلية للسحب وهو $\binom{N}{n}$ ليصبح لدينا الشكل العام لتوزيع فوق الهندسي وفقا إلى:

$$P(X = x) = \begin{cases} \frac{\binom{a}{x} \binom{N-a}{n-x}}{\binom{N}{n}}, & x = 0, 1, 2, \dots, n \\ 0, & \text{otherwise} \end{cases}$$

مثال (44)

لطلبية مؤلفة من 20 من الأشرطة الفنية كان هناك 5 معطوبة. فإذا تم اختيار أو سحب 10 من هذه الأشرطة للتفتيش والمتابعة أوجد احتمال تحديد 2 من الأشرطة معطوبة.

الحل

بالتعويض سنرى أن $x = 2$, $n = 10$, $a = 5$, $N = 20$

$$P(X = 2) = \frac{\binom{5}{2} \binom{15}{8}}{\binom{20}{10}} = 0.348$$

لذلك فإن

4-6-4 التوزيع الهندسي Geometric distribution:

ويخص هذا التوزيع الحالات التي نود فيها الحصول على عدد من المحاولات للوصول على أول نجاح للمحاولة. بمعنى أن عدد المحاولات التي تم الحديث عنها في تجارب ذي الحدين سيكون غير محددًا n is not fixed وبذلك فإن أول نجاح سيتم في المحاولة x مسبقًا بعدد من المحاولات الفاشلة بالعدد $x-1$ وإذا تم استخدام p كاحتمال نجاح أي محاولة فإن حساب الاحتمالات بهذه الصيغة سيكون حسب:

$$P(X = x) = \begin{cases} p(1-p)^{x-1}, & x = 1, 2, 3, 4, \dots \\ 0, & \text{otherwise} \end{cases}$$

ما يسمى بالتوزيع الهندسي وأن $\mu = \frac{1}{p}$

مثال
(45)

إذا كان احتمال أن أحد الأجهزة سيكون معطوبًا هو 0.05 أوجد احتمال أن الجهاز السادس هو الجهاز المعطوب.

الحل

بالتعويض لدينا: $x = 6, p = 0.05$

لذلك فإن:

$$\begin{aligned} P(x = 6) &= 0.05(0.95)^5 \\ &= 0.039 \end{aligned}$$

4-6-5 التوزيع المتكافئ Discrete Uniform Distribution:

إذا كان للوحدات المختلفة نفس الاحتمال والذي يساوي $1/n$ ، حيث أن n يمثل

عدد الوحدات فإن التوزيع يسمى توزيع منتظم. وبذلك فإن دالة التوزيع الاحتمالية ستكون:

$$P(X = x) = \begin{cases} \frac{1}{n}, & x = 1, 2, 3, 4, \dots, n \\ 0, & \text{otherwise} \end{cases}$$

مثال
(46)

لتجربة رمي حجر نرد منتظم اكتب التوزيع الاحتمالي للمتغير X والذي يمثل الرقم الظاهر على وجه الحجر.

الحل

بما أن للحجر ستة اوجه منتظمة لذلك فإن احتمال الحصول على أي من هذه الأوجه هو $1/6$ وبذلك فإن:

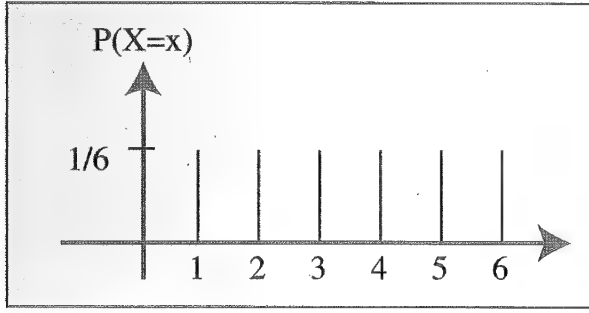
$$P(X = x) = \begin{cases} \frac{1}{6}, & x = 1, 2, 3, 4, 5, 6 \\ 0, & \text{otherwise} \end{cases}$$

ويمكن كتابة التوزيع بشكل جدول كآلاتي:

$X:$ 1, 2, 3, 4, 5, 6

1/6 1/6 1/6 1/6 1/6 1/6

أما رسم التوزيع فيتمثل بالشكل (18).



الشكل (18) لتوزيع الاحتمالات

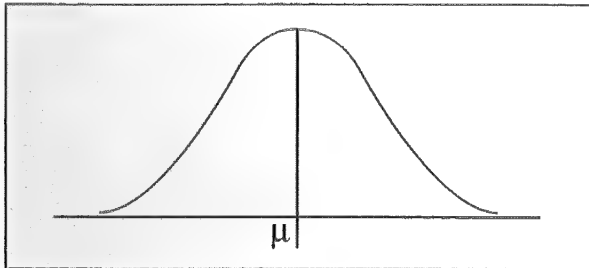
4-7 أمثلة التوزيعات المستمرة:

وكما أشرنا سابقا فإن للتوزيعات أهمية تطبيقية كبيرة ولذلك فيمكن الاستعانة بعدد من التوزيعات المستمرة والتي تساعدنا في تحديد شكل الدالة وحساب الاحتمالات المناسبة وسيتم عرض هذه التوزيعات بالشكل التالي:

1-4-7-1 التوزيع الطبيعي The Normal Distribution

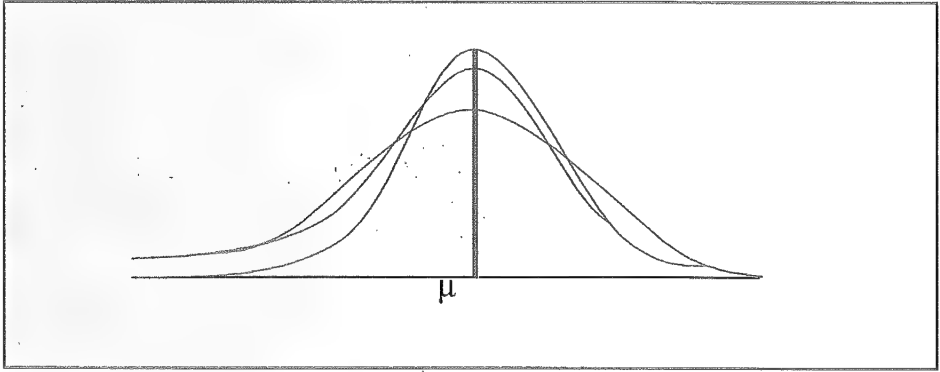
ويمكن القول بأن هذا التوزيع هو أهم التوزيعات ومجرد ذكر اسمه يعني الانطباع لذلك فهو يمثل اغلب الظواهر الطبيعية والاقتصادية وغيرها ومن خصائصه:

- 1- التماثل Symmetry ويأخذ شكل الجرس Bell shape من أهم خصائص التوزيع الطبيعي التماثل بمعنى أن قيمة الوسط الحسابي والوسيط والمنوال متساوية وان الشكل العام لهذا التوزيع التماثل هو الجرسى Bell shape او ما يسمى بالقبعة المكسيكية Mexican Hat والشكل (19) يوضح ذلك:



الشكل (19)

2- هو ليس منفردا بل مجموعة من التوزيعات Family of Distributions تسمى توزيعات طبيعية لها نفس الخصائص العامة ولكن تختلف عن بعضها في قيمة الوسط او مقدار التشتت او كليهما والشكل (20) يوضح ذلك .



الشكل (20)

3- صيغة دالته الاحتمالية هي:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \text{Exp} \left\{ -\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right\}$$

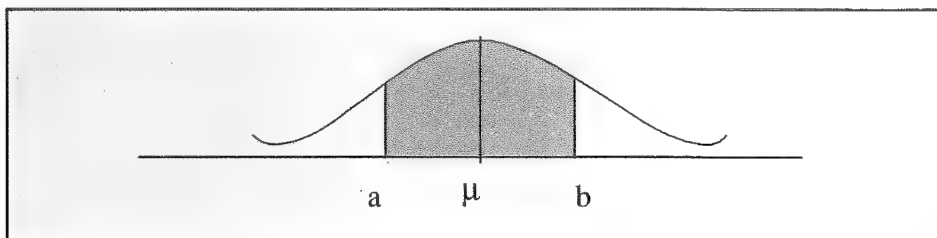
حيث أن $x \in \mathbb{R}$ وأن $\mu \in \mathbb{R}$ أما $\sigma^2 > 0$.

ويمكن أن يرمز لهذا التوزيع بالرمز $X \sim N(\mu, \sigma^2)$ أي أن المتغير X يتوزع توزيعا طبيعيا بالوسط μ و تباين σ^2 .

4- المساحة تحت المنحنى الطبيعي تساوي واحد تقسم إلى نصفين متساويين عند قيمة \bar{X} وعن إيجاد الاحتمالات الخاصة بالتوزيع الطبيعي فنستخدم:

$$P(a < x \leq b) = \int_a^b f(x) dx$$

ويمكن أن تتمثل بالمساحة تحت المنحنى الطبيعي بين النقطتين a و b وأن الشكل (21) يوضح ذلك :



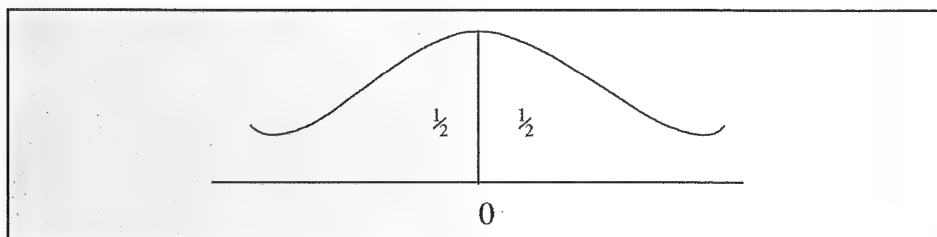
الشكل (21)

هناك جداول خاصة تسمى بجداول التوزيع الطبيعي المعياري Standard Normal Distribution لحساب تلك الاحتمالات وسيتم التعرف على كيفية استخدام تلك الجداول بعد تعريف التوزيع الطبيعي للمعياري كالآتي:

4-7-2 التوزيع الطبيعي المعياري The Standard Normal Distribution:

تم تعريف القيمة المعيارية والتي يرمز لها بالرمز Z سابقا على إنها $Z = \frac{x - \mu}{\sigma}$

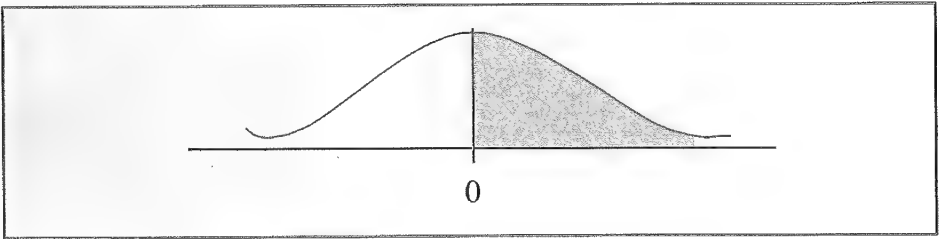
وبذلك فإن Z تتوزع توزيعا طبيعيا معياريا إذا كان X يتوزع توزيعا طبيعيا بالوسط μ والتباين σ^2 أي أن الانحراف المعياري σ ونكتب عندئذ $Z \sim N(0, 1)$ أي أن Z تتوزع توزيعا طبيعيا بالوسط صفر وتباين واحد وبذلك يمكن اعتباره حالة خاصة من التوزيعات الطبيعية السابقة الذكر وإن رسم التوزيع الطبيعي المعياري سيكون بالشكل (22).



الشكل (22)

استخدام جداول التوزيع الطبيعي:

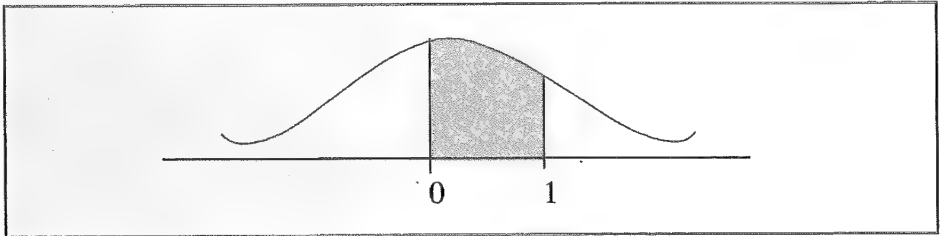
يلاحظ من جداول التوزيع الطبيعي الموجودة ضمن الملاحق أن هذه الجداول تمكّننا من إيجاد الاحتمالات المناظرة إذا أعطيت لنا قيم Z وكذلك يمكن إيجاد قيم Z المناظرة إذا أعطيت لنا الاحتمالات. في البداية نقول بأن هيكل الجدول هو أن قيم Z تكون على هامش الجدول أما الاحتمالات فتكون متن أو محتوى الجدول ولذلك يلاحظ بأن القيم الموجودة داخل الجدول تبدأ من 0.000 وتنتهي ب 0.4990 وبذلك فإن الاحتمالات تمثل نصف المساحة تحت المنحنى والشكل (23) يوضح ذلك.



الشكل (23)

أما عن قيم Z الواقعة في الهامشين فهي للقيم الموجبة وباستخدام خاصية التماثل يمكن قراءة القيم السالبة فمثلا لإيجاد $P(0 < Z < 1)$ فإننا ندخل الجدول وننزل عموديا إلى الأسفل في هامش قيم Z إلى حد الرقم 1 ثم ندخل أفقيا ضمن ذلك السطر لنقف عن أول قيمة من الاحتمالات والتي تقابل الرقم 0.3413 وذلك لأن الرقم 1 عدد صحيح ولا يوجد هناك أي مرتبة أخرى.

وبذلك فإن $P(0 < Z < 1) = 0.3413$ من الجدول ويمكن توضيحها بالشكل (24).



الشكل (24)

أما لحساب $P(0 < Z < 2.27)$ فإننا وبنفس الأسلوب السابق ننزل عموديا إلى الأسفل في قيم Z إلى حد الرقم 2.2 ثم ندخل أفقيا لنقف عند النقطة المناظرة للقيمة 0.07 في أعلى العمود ونقرأ نقطة تقاطع الصف والعمود المعينين وبذلك فإن الاحتمال هو 0.4884 وهكذا لبقية القراءات.

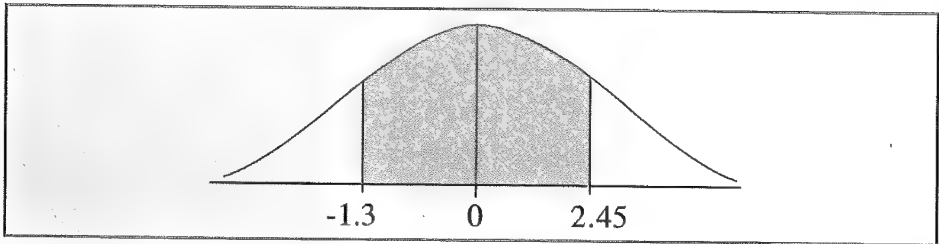
وفيما يلي سنقدم تطبيقات أخرى لقراءة القيم الجدولية كآلاتي:

مثال (47)

أوجد $P(-1.3 < Z < 2.45)$

الحل

لإيجاد هذا الاحتمال يمكن الاستعانة بالشكل (25) الذي يمثل المساحة المراد قراءتها.



الشكل (25)

وبذلك يتضح من الشكل أن المساحة المراد حسابها هي جمع المساحتين من النقطة صفر إلى الرقم 2.45 ومن النقطة صفر إلى الرقم 1.3 مرة أخرى وبذلك فإن المساحة المطلوبة تكون جمع المساحتين أي أن الاحتمال المطلوب سيكون جمع الاحتمالين 0.4032 و 0.4929 لذلك فإن :

$$P(-1.3 < Z < 2.45) = 0.4032 + 0.4929 = 0.8961$$

مثال (48)

إذا علمت أن أطوال مجموعة من الطلبة تتوزع طبيعياً بالوسط 160 سم وانحراف معياري 5 سم. أوجد احتمال اختيار طالب يكون طوله ما بين 155 سم و165 سم.

الحل

حل هذا المثال علينا أولاً بتحويل التوزيع الطبيعي العادي إلى التوزيع الطبيعي المعياري باستخدام صيغة القيم المعياري السابق ذكرها ومن ثم يتم حساب الاحتمال و يتم ذلك كآلاتي:

$$\begin{aligned}
 P(155 < X < 165) &= P\left(\frac{155 - 160}{5} < \frac{X - \mu}{\sigma} < \frac{165 - 160}{5}\right) \\
 &= \begin{cases} P(-1 < Z < 1) \\ 2P(0 < Z < 1) \end{cases} \\
 &= 2(0.3413) \\
 &= 0.6826
 \end{aligned}$$

مثال (49)

أوجد قيمة Z بحيث أن $P(0 < Z < z_0) = 0.3470$.

الحل

حل هذا المثال علينا الدخول للجدول بطريقة معاكسة لما تم عمله سابقاً. فهنا علينا الدخول إلى متن الجدول بحثاً عن الاحتمال أو المساحة تحت المنحنى بالمقدار 0.4370 فإذا تم إيجادها فنقف عندها، أما إذا لم يتم إيجادها بالضبط فنحاول إيجاد أقرب قيمة لها (كما سيتم ملاحظته بالمثال التالي). وعند تلك النقطة نقرأ قيمة Z من هامشي الجدول.

لذلك فمن داخل الجدول نقف أو نضع الإصبع على الرقم 0, 734 ويكون عند نقطة تقاطع الصف الذي يحمل قيمة 1, 5 والعمود الذي يحمل قيمة 0, 30 لذلك فإن z_0 المقابلة لهذا الاحتمال ستكون : $z_0 = 1.5 + 0.03 = 1.53$

مثال (50)

أوجد قيمة z_0 بحيث أن $P(0 < Z < Z_0) = 0.2400$.

الحل

بما أن القيمة 0.2400 ليست موجودة فنحاول احتساب اقرب قيمة لها والتي تقع بين القيمتين 0.2389 المقابلة إلى z_0 بالمقدار 0.64 والقيمة 0.2422 المقابلة إلى z_0 بالمقدار 0.65 لذلك فإن z_0 المطلوبة ستقع بين القيمتين 0.64 و 0.65 ، وبالطريقة المناسبة نجد أن :

$$z_0 = 0.645$$

ملاحظة:

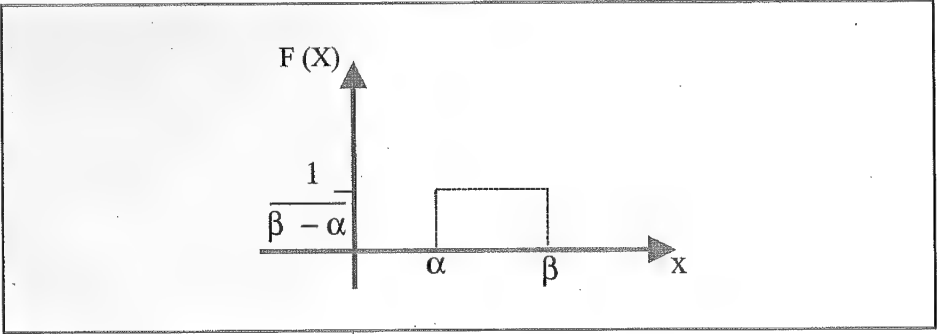
يمكن القول بأن التوزيع الطبيعي هو التوزيع التقريبي لجميع التوزيعات الأخرى عندما يكون حجم العينة كبيرا، أما عن الحجم المناسب للقول بأن الحجم كبيراً فيعتمد على أسس كثيرة ولكن يمكن القول بأن حجم عينة $n \geq 30$ يكون مناسباً لاعتبار أن حجم العينة كبيراً أما عندما $n < 30$ فيمكن اعتباره صغيراً، وأخيراً فإن تحويل أي توزيع إلى التوزيع الطبيعي يتم باستخدام صيغة القيم المعيارية Z بعد تحديد قيمة الوسط وقيمة الانحراف المعياري المناسب.

3-4-7 التوزيع المنتظم Uniform distribution:

للتوزيع معلمتان هما α و β والدالة الاحتمالية هي :

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \alpha < x < \beta \\ 0, & \text{otherwise} \end{cases}$$

وبذلك إن هذا التوزيع له شكل مستطيل عند الرسم كما يظهر من الشكل (26).



الشكل (26)

ولذلك يمكن أن يطلق عليه اسم التوزيع المستطيلي وبمعنى المنتظم نعني أنه ولجميع نقاط هذا التوزيع فإن دالة الكثافة النسبية متساوية وأن المساحة لهذا المستطيل تساوي الواحد.

أما الوسط والتباين فهما : $\mu = \frac{\alpha + \beta}{2}$ و $\sigma^2 = \frac{(\beta - \alpha)^2}{12}$ على التوالي.

4-7-4 توزيع كاما The Gamma distribution:

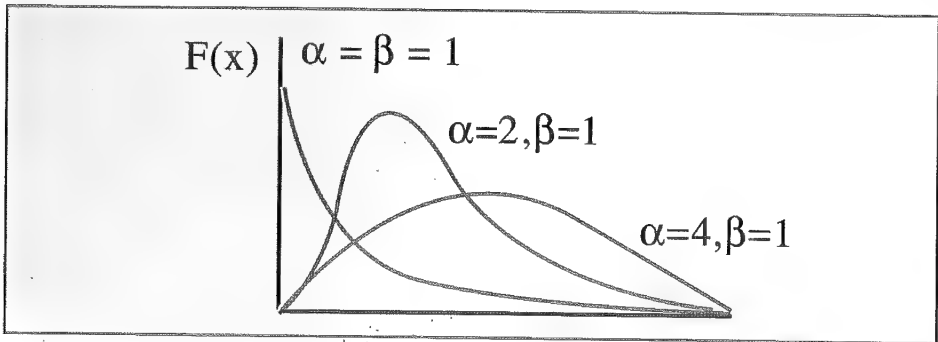
توزيع مهم و مفيد و يتضمن العديد من التوزيعات الأخرى كحالات خاصة من هذا التوزيع. ولهذا التوزيع الدالة الاحتمالية التالية:

$$\sigma^2 = \frac{(\beta - \alpha)^2}{12} \quad \mu = \frac{\alpha + \beta}{2}$$

حيث أن المعلمتان α و β تتحددان بالقيم $\alpha > 0$ و $\beta > 0$ وان :

$$\begin{aligned} \Gamma(\alpha) &= \int_0^{\infty} x^{\alpha-1} e^{-x} dx \\ &= (\alpha - 1) \Gamma(\alpha - 1) \end{aligned}$$

للقيم $\alpha > 1$ وأن تكون أعداد صحيحة موجبة لدينا $\Gamma(\alpha) = (\alpha - 1) \Gamma(\alpha - 1)$ ، الشكل (27) التالي يبين عدد من توزيعات جاما بالاستناد لقيم معينة من المؤشرين α و β



الشكل (27)

أما عن وسط وتباين هذا التوزيع فهما $\sigma^2 = \alpha\beta^2$, $\mu = \alpha\beta$ على التوالي. عندما $\alpha = 1$ فإن توزيع كاما سيمثل حالة خاصة وعندئذ يسمى بالتوزيع الأسّي Exponential Distribution بدالة الكثافة الاحتمالية التالية:

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & otherwise \end{cases}$$

ويعتمد على مؤشر واحد هو $\beta > 0$

4-7-5 توزيع بيتا The Beta Distribution:

$$f(x) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, & 0 < x < 1 \\ 0, & Otherwise \end{cases}$$

معتمدا على مؤشرين هما $\alpha > 0$ و $\beta > 0$.

وعن وسط وتباين هذا التوزيع فهما على التوالي:

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}, \quad \mu = \frac{\alpha}{\alpha + \beta}$$

أسئلة

الفصل الرابع

1- ليكن $A_1 = \{0, 1, 2\}$ و $A_2 = \{2, 3, 4\}$ أوجد $A_1 \cup A_2$ و $A_1 \cap A_2$.

2- ليكن $A_1 = \{x : x = 0, 1, \dots, 10\}$ و $A_2 = \{x : x = 8, 9, 10, 11\}$ أوجد $A_1 \cup A_2$ و $A_1 \cap A_2$.

3- ليكن $A_k = \{x : 0 \leq x \leq \frac{1}{k}\}$ و $k = 1, 2, 3, \dots$ أوجد $A_1 \cup A_2 \cup A_3 \dots$ و $A_1 \cap A_2 \cap A_3 \dots$.

4- ليكن $A_k = \{x : \frac{1}{k+1} \leq x \leq 1\}$ و $k = 1, 2, 3, \dots$ أوجد $A_1 \cup A_2 \cup A_3 \dots$ و $A_1 \cap A_2 \cap A_3 \dots$.

5- ليكن $A_1 = \{(0,0), (0,1), (1,1)\}$ و $A_2 = \{(1,1), (1,2), (2,1)\}$ أوجد $A_1 \cup A_2$ و $A_1 \cap A_2$.

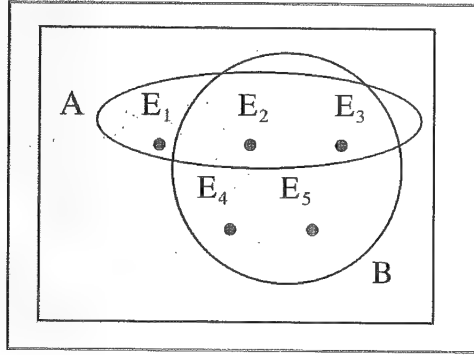
6- ليكن $A_1 = \{(x, y) : 0 < x < 2, 0 < y < 2\}$ و $A_2 = \{(x, y) : 1 < x < 3, 1 < y < 3\}$ أوجد $A_1 \cup A_2$ و $A_1 \cap A_2$.

7- 500 قرض لشراء السيارات مقدم من أحد البنوك للسنة الماضية ظهر ضمن الجدول التالي:

كمية القرض :	اقل من 1000	1000- 4000	6000 فاكتر
العدد :	27	99	298
			76

أوجد:

- احتمال أن القروض المختارة سيكون 6000 فأكثر.
 - احتمال أن أحد القروض المختارة سيكون أقل من 4000.
 8- إذا علمت أن فضاء عينة يتألف من خمسة أحداث بسيطة بالشكل:



وأن $P(E_1) = 0.1$ و $P(E_2) = 0.1$ و $P(E_3) = 0.3$ و $P(E_4) = 0.4$

وأن $P(E_5) = 0.1$

أوجد $P(A)$ و $P(B)$ و $P(A \cap B)$ و $P(A \cup B)$ و $P(A^c)$ و $P(B^c)$.

- 9- إذا علمت أن فضاء عينة يتألف من ثلاث أحداث متنافية هي A, B, C وأن $P(A) = 0.4$ و $P(B) = 0.25$ أوجد $P(C)$ ثم أوجد:

$P(A \cap B)$ و $P(A \cup B)$ و $P(A|B)$.

- 10- الجدول الثنائي التالي يبين تصنيف مجموعة من الأشخاص حسب العمر والدخل كالاتي:

الدخل \ العمر	أقل من 2000	2000 -	5000 فأكثر
أقل من ٢٥	950	1000	50
٢٥ -	450	2050	1500
٤٥ فأكثر	50	950	1000

عرف الأحداث التالية:

الحدث A أن يكون عمر الشخص بين اقل من 25.

الحدث B أن يكون عمر الشخص بين 25 و 45.

الحدث C أن يكون عمر الشخص 45 فاكثر.

الحدث D أن يكون دخل الشخص اقل من 2000.

الحدث E أن يكون دخل الشخص بين 2000 و 5000.

الحدث F أن يكون دخل الشخص 5000 فاكثر.

أوجد $P(A)$ و $P(B)$ و $P(C)$ و $P(D)$ و $P(E)$ و $P(F)$ و $P(A \cap B)$ و $P(A \cup B)$

و $P(C \cap F)$ و $P(C \cup F)$ و $P(A^c)$ و $P(A^c \cap C)$ و $P(A^c \cup C)$

و $P(B^c \cap E)$ و $P(B^c \cup E)$.

11 - الجدول الثنائي التالي يمثل تصنيف مجموعة من المتسوقين كالاتي:

	يتسوق من المحل	لا يتسوق من المحل
شاهد الإعلان	100	25
لم يشاهد الإعلان	25	50

أوجد:

- احتمال أن شخص اختير عشوائيا قد شاهد الإعلان.

- احتمال أن شخص اختير عشوائيا يتسوق من المحل.

- احتمال أن شخص اختير عشوائيا شاهد الإعلان ويتسوق من المحل.

- احتمال أن شخص اختير عشوائيا شاهد الإعلان أو يتسوق من المحل.

- احتمال أن شخص اختير عشوائيا يتسوق من المحل علما بأنه شاهد الإعلان.

12 - لتجربة سحب ورقتان من مجموعة من أوراق اللعب المؤلفة من 52 ورقة.

افترض أن C_1 يمثل الحصول على ورقة قلب وأن C_2 يمثل الحصول على ملك.

أوجد $P(C_1)$ و $P(C_2)$ و $P(C_1 \cap C_2)$ و $P(C_1 \cup C_2)$.

13 - وعاء يحتوي على 16 قرص منها 6 حمراء و 7 بيضاء و 3 زرقاء. وبافتراض

أننا نود سحب 4 أقراص من هذا الوعاء, أوجد احتمال أن الأقراص الأربعة

حمراء: أ) بدون إرجاع. ب) بالإرجاع.

14 - صندوق يحتوي على 8 كرات منها 3 كرات حمراء والباقية زرقاء. وبافتراض أننا نود سحب كرتين، أوجد احتمال أن الكرة الأولى حمراء والثانية زرقاء أ) بدون إرجاع ب) بالإرجاع.

$$15 - \text{ليكن} \quad P(X = x) = \begin{cases} \frac{x}{15}, & x = 1, 2, 3, 4, 5 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{أوجد } P(1 \leq X \leq 2) , P\left(\frac{1}{2} < X \leq \frac{5}{2}\right) P(X = 1 \text{ OR } 2)$$

$$16 - \text{ليكن} \quad f(x) = \begin{cases} cx, & x = 1, 2, 3, 4, 5, 6 \\ 0, & \text{otherwise} \end{cases}$$

أوجد قيمة c بحيث أن f هي دالة احتمالية.

$$17 - \text{ليكن} \quad f(x) = \begin{cases} c\left(\frac{2}{3}\right)^x, & x = 1, 2, 3, \dots \\ 0, & \text{otherwise} \end{cases}$$

أوجد قيمة c بحيث أن f هي دالة احتمالية.

18 - ليكن $z \sim N(0, 1)$ أوجد:

$$.P(-1.52 < z < 0) . P(0 < z < 2.78) . P(0 < z < 2.70)$$

$$. P(z < 2) . P(-1.23 < z < 0.75) . P(-0.95 < z < 0.95)$$

$$. P(z < -2.3) . P(z < -1) . P(z < 2.5)$$

19 - ليكن $x \sim N(100, 25)$ أوجد:

$$. P(x < 97) . P(x > 90) . P(x < 103.5) . P(100 < x < 110)$$

$$. P(95 < x < 105) . P(x > 105)$$

20- إذا علمت أن أطوال مجموعة من الطلبة تتوزع توزيعاً طبيعياً بوسط 165 سم وتباين 100 سم أوجد احتمال أن طالبا اختير عشوائياً وكان طوله:

- بين 160 سم و 170 سم.

- أطول من 163 سم.

- أطول من 170 سم.

- أقصر من 160 سم.

21- إذا علمت أن مبيعات مصنع معين تتوزع توزيعاً طبيعياً بوسط 500000 دينار وانحراف معياري 50000 دينار. أوجد احتمال أن مصنع اختير عشوائياً وكانت مبيعاته:

- أكثر من 600000 دينار.

- أقل من 400000 دينار.

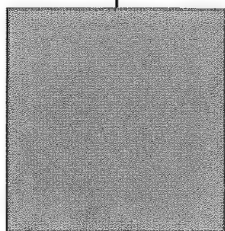
22- إذا علمت أن درجات طلبة تتوزع توزيعاً طبيعياً بوسط 70 وتباين 25 أوجد احتمال أن طالب اختير عشوائياً وكانت درجته:

- بين 60 و 80.

- أقل من 50.

- أكثر من 90.

الاحياء



5

الفصل الخامس

توزيعات
المعاينة

**Sampling
Distributions**

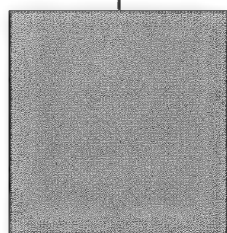
5-1 نظرية المعاينة

5-2 أنواع العينات

5-3 توزيع المعاينة

الفصل الخامس

5



الفصل الخامس

توزيعات المعاينة

Sampling Distribution

5

5-1 نظرية المعاينة Sampling Theory :

ذكرنا سابقاً أن علم الإحصاء يمكن تقسيمه إلى قسمين ؛ الإحصاء الوصفي Descriptive statistics والإحصاء الاستدلالي Statistical Inference والذي يمكن الباحث من التوصل إلى خواص المجتمعات عن طريق العينات أو المعاينة حيث أن المجتمع هو عبارة عن جميع المفردات التي يمكن أن يأخذها المتغير، أما العينة فهي جزء من المجتمع لأن دراسة المجتمع تكون في بعض الأحيان مستحيلة أو مكلفة وتحتاج إلى جهد و وقت طويل.

العينة Sample:

وهي مجموعة جزئية من مفردات المجتمع الإحصائي التي يتم جمعها بحيث تكون ممثلة لها إذا هي عنصر من عناصر أو فرد من أفراد المجتمع الذي ندرسه.

الاختيار العشوائي Random Selection:

وهي طريقة اختيار مفردات العينة من المجتمع الإحصائي المطلوب إجراء دراسة عليه ويقال أن الاختيار تم عشوائياً إذا كان لكل مفردة من مفردات المجتمع لها نفس فرصة " الاختيار ".

الإحصاءة Statistic:

وهي عبارة عن كل قيمة تحسب في العينة أو بعبارة أخرى هي عبارة عن متغير عشوائي قيمة هذا المتغير تعتمد على العينة Sample ، فالإحصاءة هي قيمة متغيرة لأنها تختلف من عينة إلى أخرى داخل المجتمع الواحد .

أن التغير في قيمة الإحصاء يعتمد على حجم المجتمع و حجم العينة و يعتمد أيضا على طريقة اختيار العينة.

5-2 أنواع العينات:

5-2-1 العينة العشوائية البسيطة Simple Random Sample:

وهي العينة التي يتم اختيار مفرداتها باستخدام طريقة الاختيار العشوائي بحيث يكون لجميع وحدات المعاينة sample units في المجتمع نفس الفرصة أو الاحتمال في الاختيار ويتم هذا عن طريق حجم المجتمع فإذا كان حجم المجتمع هو N فإن احتمال اختيار أي مفردة منه هو $\frac{1}{N}$ ويمكن اختيار العينة بالإرجاع Replacement أي إرجاع وحدة المعاينة المسحوبة قبل سحب الوحدة التي تليها أو اختيار العينة بدون إرجاع Sampling without Replacement أي بدون إرجاع وحدة المعاينة المسحوبة قبل سحب العينة التي تليها.

وان من أبسط أنواع اختيار عينة عشوائية حجمها n هي بأن تسجل وحدة المعاينة لجميع المجتمع على بطاقات متشابهة تماما تعطي تسلسل بحيث أن كل بطاقة تأخذ رقما واحدا ثم تخلط جيدا ثم يتم سحب بطاقة وراء الأخرى مع إرجاع البطاقة بعد سحبها وتدوين الرقم الموجود عليها ومن ثم يتم التوصل إلى عدد مفردات مساوية لحجم العينة في حالة اختيار العينة مع الإرجاع.

أما إذا كان حجم المجتمع كبيرا جدا فيستحسن استعمال جداول الأعداد العشوائية والذي يظهر في الملحق وهو جدول مكون من مجموعات من الأعمدة كل منها مكون من أعداد قد يكون العدد مكون من رقم أو رقمين فإذا فرضنا أن المجتمع مكون من 560 مفردة أردنا اختيار عينة عشوائية بنسبة 10% من المجتمع فذلك يعني أن عدد مفردات هذه العينة يساوي 65 مفردة يتم اختيارها كما يلي:

بطريقة الاختيار العشوائي نحدد رقم الصف والعمود الذي نبدأ به من العدد الموجود و تقاطعهما ليكون أول مفردة، ثم نستمر في اخذ الأعداد المتتالية على أساس التتابع الرأسي حتى نحصل على مجموعة من الأعداد تساوي حجم العينة.

5-2-2 المعاينة المنتظمة Systematic Sample:

هي طريقة اختيار عينة منتظمة وذلك عن طريق اختيار وحدة المعاينة المرقمة K (والتي تسمى نسبة المعاينة ، وتمثل حجم المجتمع إلى حجم العينة) ثم اختيار رقم عشوائياً بين 1 و K ليمثل رقم الوحدة الأولى ثم إضافة K ومضاعفاتها على رقم الوحدة الأولى إلى أن تكتمل حجم العينة.

ففي المثال السابق كان حجم المجتمع 560، والعينة 56 فإن نسبة المعاينة:

$$\text{نسبة المعاينة } K = \frac{\text{حجم المجتمع } (N)}{\text{حجم العينة } (n)} = \frac{560}{56} = 10$$

لذا نقسم مفردات المجتمع إلى مجموعات عددها يساوي حجم العينة وعدد المفردات داخل كل مجموعة يساوي نسبة المعاينة وبإضافة نسبة المعاينة إلى رقم أول مفردة ثم التالية لها إلى أن نحصل على حجم العينة المطلوب .

ثم نختار رقماً عشوائياً بين 1 و 10 وليكن 5 فهذا يكون رقم الوحدة الأولى ثم نضيف 10 ومضاعفاتها إلى الوحدة الأولى لنحصل على الوحدات 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55 وهكذا إلى أن يكتمل حجم العينة المؤلفة من 56 وحدة معاينة.

5-2-3 العينة العشوائية الطبقية Stratified Random Sample:

وهي طريقة اختيار عينة طبقية عن طريق تقسيم المجتمع إلى مجموعات غير متداخلة أي متجانسة تسمى طبقات Strata ثم اختيار عينة عشوائية فرعية من كل طبقة بحيث أن العينات الفرعية مجتمعة تكون العينة الطبقية العشوائية ، وعادة ما أن يكون حجم العينة العشوائية الفرعية البسيطة متناسباً مع حجم الطبقة وهذه الطريقة تسمى التخصيص النسبي Proportional وأحياناً يكون حجم العينة الفرعية متساوياً لجميع طبقات المجتمع.

أما كيفية اختيار العينة الطبقية عشوائياً، يجب تعيين الطبقات أولاً بوضوح وبعد ذلك وضع كل وحدة معاينة من المجتمع في الطبقة الملائمة، وبالتالي تحديد حجم كل

طبقة بعد ذلك نحدد حجم العينة من كل طبقة ومن ثم نختار عينة عشوائية بسيطة من كل طبقة.

أما أسلوب تقسيم العينة على الطبقات فيتم باتباع طريقة النسبة و هي الحصول على عينة تمثل المجتمع فإذا كان لدينا K من الطبقات فاختيار عينة عشوائية حجمها n من المجتمع بأسلوب النسبة لو فرضنا أردنا حساب n_k من الطبقات فإن n_1, n_2, \dots, n_k تكون حجم العينة من كل طبقة فإذا حجم العينة الكلي $n_1 + n_2 + \dots + n_k$ وهنا يجب تحديد حجم العينة في أي طبقة بحيث يتناسب مع حجم الطبقة فذا كان لدينا N حجم المجتمع أما N_1, N_2, \dots, N_k هو حجم الطبقات. فلاحساب حجم العينة من كل طبقة يكون كما يلي :

$$n_k = N_k \left(\frac{n}{N} \right) \quad \text{حجم العينة في كل طبقة}$$

مثال (1)

إذا كان عدد المسجلين من طلاب السنة الأولى في جامعة عمان الأهلية حسب الكليات:

الآداب: 1200 طالباً والعلوم: 2000 طالباً والاقتصاد: 1400 طالباً والعلوم الطبية: 800 طالباً.

المطلوب إيجاد عينة حجمها 20% من المجتمع.

$$\text{حجم المجتمع} = 1200 + 2000 + 1400 + 800 = 5400$$

$$\text{وأن حجم العينة الكلي} = 5400 \times 20\% = 1080$$

وبذلك فإن حجم العينة الأولى من كلية الآداب هي: $1200 \times 20\% = 240$

حجم العينة من كلية العلوم هي: $2000 \times 20\% = 400$

حجم العينة من كلية الاقتصاد هي: $1400 \times 20\% = 280$

حجم العينة من كلية العلوم الطبية هي: $800 \times 20\% = 160$

4-2-5- العينة المشوائية المتعددة المراحل Multi-Stage Random Sample:

وهي طريقة لاختيار على مراحل متعددة واختيار مفرداتها في كل مرحلة. فإذا كانت وحدات المجتمع مقسمة إلى أقسام فإننا في المرحلة الأولى نختار عشوائيا عينة من هذه الأقسام وفي المرحلة الثانية نختار عشوائيا من العينة التي اختبرت في المرحلة الأولى وهكذا حتى نحصل على حجم العينة المقرر.

مثال على ذلك عند اختيار عينة من المحافظات في المرحلة الأولى واختيار عينة من المراكز ضمن المحافظات كمرحلة ثانية وهكذا

والأسلوب الأمثل لتحديد العينة من مجتمع محدود كما يلي:

$$n = \frac{p(1-p)}{\frac{p(1-p)}{N} + \frac{\alpha^2}{Z^2}}$$

حيث p = نسبة معينة

n = حجم المجتمع

α = مستوى (المعنوية)

Z = الدرجة المعيارية المقابلة إلى معامل الثقة $(1-\alpha)$ يتم حسابها من جدول Z الطبيعي.

4-3-5 توزيع المعاينة Sampling Distribution:

يخضع المجتمع الذي تؤخذ منه العينة لتوزيع معين وهو توزيع المجتمع الاحتمالي لمتغير عشوائي يمثل وحدات ذلك المجتمع. وان التوزيع الاحتمالي للإحصاء يدعى بتوزيع المعاينة لتلك الإحصاءة والمتمثل عادة بثوابت تعين هذا التوزيع تماما وتسمى معالمات Parameters. فمثلا إذا كان المجتمع يخضع لتوزيع طبيعي فإن المعلمين هما الوسط الحسابي μ والانحراف المعياري σ . فإذا كانت معلومة فيمكن عندئذ إيجاد جميع الاحتمالات المتوقعة لهذا المجتمع. وكذلك الحال إن كان

المجتمع يخضع لتوزيع ذو حدين فإن المعلمة هي احتمال النجاح p . فإذا علمت p فإننا نستطيع معرفة توزيعه أي يمكن تحديد مجتمعه. أي أن دالة الكثافة الاحتمالية للمتغير x والذي يمثل أي فرد من أفراد ذلك المجتمع يكون قد تحدد تماما. وبعد ذلك يمكن حساب بعض المقاييس عن هذه العينة مثل الوسط الحسابي للعينة الواحدة هو (\bar{X}) و يسمى بإحصاءة العينة وهذه القيمة ربما تتغير من عينة إلى أخرى، وإن قيمة الوسط كما ذكرنا فإنها تتغير من عينة إلى أخرى لذا فالمتغير العشوائي هنا هو (\bar{X}) أي أن قيمة هذا الإحصاءة \bar{X} تتغير بتغير العينة. ولهذا يسمى التوزيع الاحتمالي لإحصاءة العينة. ومنها يمكن التوصل إلى تعريف الانحراف القياسي لتوزيع المعاينة والذي يدعى بالخطأ القياس للإحصاءة.

نظرية: إذا كان (X) يخضع لتوزيع وسطه μ وتباينه σ^2 و كان (\bar{X}) يمثل الوسط الحسابي لعينة حجمها n والمسحوبة من هذا المجتمع فإن القيمة المتوقعة لهذا الوسط هو $(\mu_{\bar{X}} = \mu)$ أي أن الوسط الحساب لـ (\bar{X}) هو الوسط الحسابي لجميع الأوساط الحسابية للعينات التي سحبت منها هذه العينات. أما تباين (\bar{X}) هو:

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

أي أن تباين هذه العينات يعتمد على تباين المجتمع وعلى حجم العينة وهو بذلك أقل من تباين المجتمع. وبالتالي كلما كبر حجم العينة قل مقدار الخطأ القياسي للوسط الحسابي (\bar{X}) ونقرب وسط تلك العينة من الوسط الحسابي للمجتمع لذا يمكن استخدام تقدير الوسط الحسابي كتقدير لـ μ ، ويجب توفر شرط الإرجاع.

1-3-5 توزيع المعاينة للوسط الحسابي من مجتمع طبيعي:

Sampling Distribution for the Mean of Normal population

إذا سحبت عينة عشوائية حجمها n من مجتمع كبير له وسط حسابي 5 وتباين معلوم σ^2 فإن توزيع المعاينة للوسط الحسابي (\bar{X}) يقترب من التوزيع الطبيعي بوسط حسابي: $\mu_{\bar{X}} = \mu$

$$\sigma_x^2 = \frac{\sigma^2}{n}$$

وإن Z هي قيمة المتغير الطبيعي القياس الذي له وسط حسابي مساويا صفر وبتباين مقداره واحد.

وهذا ما يسمى بتوزيع المعاينة للوسط الحسابي (\bar{X})

مثال (2)

تخضع علامات الطلاب في مادة الإحصاء لتوزيع طبيعي بمعدل 70 وانحراف معياري 20 سحبت منه عينة عشوائية حجمها 36 طالباً، أوجد: توزيع المعاينة لهذه العينة.

ثم احسب احتمال أن يزيد معدل علامات الطلاب عن 78.

الحل

$$\mu_{\bar{x}} = \mu = 70$$

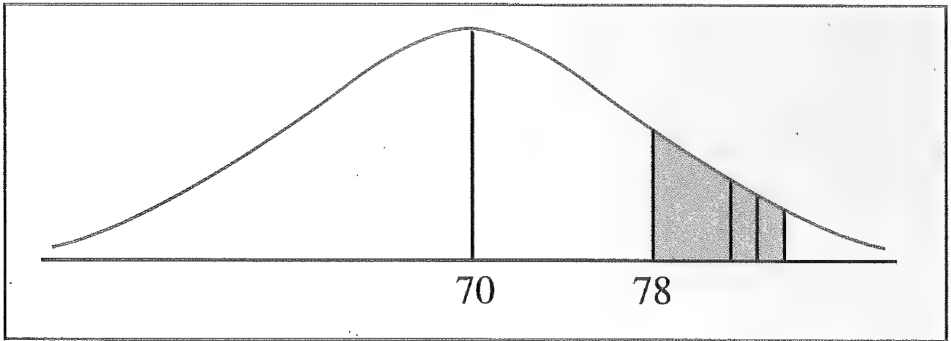
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{36}} = \frac{20}{6} = 3.3$$

وإن توزيع المعاينة هو:

$$\bar{X} \sim N(70, 10.89)$$

لإيجاد الاحتمال لدينا

$$\begin{aligned} p(\bar{X} \geq 78) &= p\left(\frac{\bar{x} - 70}{3.3} \geq \frac{78 - 70}{3.3}\right) \\ &= p(Z \geq 2.42) \end{aligned}$$



$$P(Z \geq 2.42) = 0.5 - p(0 < Z < 2.42)$$

و باستخدام جدول التوزيع الطبيعي:

$$P(Z > 2.42) = 0.5 - 0.4922 = 0.0074$$

2-3-5- توزيع المعاينة للفرق بين وسطين:

Sampling Distribution for the Difference Between Two Sample Means

نفرض أن لدينا مجتمعين الأول بوسط (μ_1) وتباينه (σ_1^2) والمجتمع الثاني وسطه الحسابي (μ_2) وتباين (σ_2^2) والمجتمعين يتبعان التوزيع الطبيعي:

نظرية:

سحبت عينة عشوائية حجمها n_1 من مجتمع طبيعي معدله (μ_1) وتباينه (σ_1^2) ، وعينة ثانية من مجتمع طبيعي أيضا معدله (μ_2) وتباينه (σ_2^2) والعينة الثانية مستقلة عن العينة الأولى فإذا كان (\bar{X}_2) يمثل الوسط الحسابي للعينة الأولى، و (\bar{X}_1) يمثل الوسط الحسابي للعينة الثانية، فإن توزيع المعاينة للفرق بين متوسطيهما $(\bar{X}_1 - \bar{X}_2)$ يتبع التوزيع الطبيعي ذا المعدل $(\mu_1 - \mu_2)$ والتباين:

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$(\bar{X}_1 - \bar{X}_2) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \quad \text{فإن توزيع المعاينة يكون:}$$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

يخضع لتوزيع طبيعي معياري بوسط مساوي إلى الصفر وانحراف معياري

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 \quad \text{يساوي واحد أي أن:}$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

مثال (3)

سحبت عينتين عشوائيتين من شركتين مختلفتين لإنتاج العدد الزراعية وكانت الأجور المدفوعة من قبل الشركتين تتبع التوزيع الطبيعي. وان معدل الأجور المدفوعة من قبل الشركة A إلى 36 عاملاً تساوي ديناراً أردنياً بانحراف معياري 36 ديناراً. أما الأجور المدفوعة من قبل الشركة B إلى 49 عاملاً تساوي 186 ديناراً بانحراف معياري 40 ديناراً. احسب احتمال أن معدل الأجور المدفوعة من قبل الشركة A لها متوسط على الأقل 60 ديناراً فاكثراً من متوسط الأجور المدفوعة للشركة B.

الحل

المعلومات المتوفرة في المثال يمكن توضيحها كما يلي:

	الشركة B	الشركة A
الوسط الحسابي	$\mu_2 = 180$	$\mu_1 = 230$
الانحراف المعياري	$\sigma_2 = 40$	$\sigma_1 = 36$
حجم العينة	$n_2 = 49$	$n_1 = 36$

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 = 230 - 180 = 50$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(36)^2}{36} + \frac{(40)^2}{49}} = 8.3$$

$$P(\bar{X}_1 - \bar{X}_2 \geq 60) = P\left\{ \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq \frac{60 - 50}{8.3} \right\}$$

$$P\left(Z \geq \frac{10}{8.3}\right) = P(Z \geq 1.204) = 0.5 - P(0 \leq Z \leq 1.204) \\ = 0.5 - 0.3925 \\ = 0.1075$$

3-3-5- توزيع المعاينة للنسب Sampling Distribution for Proportion

إذا كانت قيمة كل عنصر متمثلة بالنجاح أو الفشل فإننا نسمي كل مشاهدة من هذا المجتمع "تجربة بيرنولي" حيث أن احتمال النجاح يساوي p واحتمال الفشل يساوي q علماً بأن $p + q = 1$ وعندما نسحب عينة عشوائية حجمها n فإنه يجب إعادة التجربة n من المحاولات وبذلك فإن توزيع المعاينة للمتغير العشوائي x المتمثل بعدد النجاحات في العينات ذات حجم n يمكن أن يكون قريباً من التوزيع الطبيعي بوسط حسابي قدره $\mu = np$ وانحراف قياسي

$\sigma = \sqrt{npq}$ علماً بأن نسبة النجاح مختلفة من عينة إلى أخرى وعلى شرط ألا تكون قريبة من الصفر أو الواحد.

يمكن التعبير عن نسبة النجاح p بالمقدار: $\hat{P} = \frac{x}{n}$

حيث أن: x = عدد المحاولات (النجاحات).

n = حجم العينة.

\hat{p} = نسبة النجاح في العينة.

وكما ذكرنا أن \hat{p} تختلف من عينة إلى أخرى فإن توزيع المعاينة لـ x عدد النجاحات

في العينات ذات الحجم n يمكن أن يكون قريباً من التوزيع الطبيعي بوسط حسابي قدره:

$$\mu_{\hat{p}} = E\left(\hat{P}\right) = E\left(\frac{X}{n}\right) = \frac{np}{n} = p$$

$$\sigma_{\hat{p}}^2 = \text{var}\left(\hat{p}\right) = \text{var}\left(\frac{X}{n}\right) = \frac{npq}{n^2} = \frac{pq}{n}$$

وبتباين

على شرط أن تكون قيم p قريبة من الصفر أو الواحد.

نظرية

سحبت عينة عشوائية حجمها n من مجتمع إحصائي يخضع لتوزيع بيرنولي، أي ذات الحدين $B(n, p)$ بوسط $\mu=np$ وتباين $\sigma^2=npq$ فتوزيع المعاينة إلى (\hat{p}) نسبة النجاحات هو قريب من التوزيع الطبيعي بوسط $(\mu_{\hat{p}}=p)$ وانحراف قياسي:

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

لذا فإن القيمة المعيارية Z هي :

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

$$\hat{p} = \frac{X}{n}$$

مثال

(4)

إذا كان احتمال نجاح الطالب في إحدى المساقات هو 0.8 ، سحبت عينة عشوائية حجمها 49 طالبا أوجد احتمال $(0.7 < \hat{p} < 0.92)$.

الحل

باستخدام النظرية يمكن إيجاد قيم الاحتمال أعلاه كما يلي:

$$P(0.7 \leq \hat{p} \leq 0.92)$$

$$\begin{aligned}
& P \left\{ \frac{0.7 - 0.8}{\sqrt{\frac{(0.8) \times (0.2)}{49}}} \leq Z \leq \frac{0.92 - 0.8}{\sqrt{\frac{(0.8) \times (0.2)}{49}}} \right\} \\
& = P \left\{ \frac{-0.1}{0.06} \leq Z \leq \frac{0.12}{0.06} \right\} \\
& = P(-1.66 \leq Z \leq 2) = P(0 \leq Z \leq 2) + P(0 \leq Z \leq 1.66) \\
& = 0.4772 + 0.4515 = 0.9287
\end{aligned}$$

5-3-4 توزيع المعاينة للفرق ما بين نسبتي:

Sampling Distribution for differences between two proportions

نظرية:

سحبت عيتان عشوائيتان حجمهما n_2, n_1 من مجتمعين مستقلين يخضع الأول لتوزيع ذي الحدين

$B(n_1, p_1)$ والثانية أيضا تخضع لتوزيع ذو حدين $B(n_2, p_2)$ وان :

$$\mu_1 = n_1 p_1 \quad \text{و} \quad \sigma^2_1 = n_1 p_1 q_1$$

$$\mu_2 = n_2 p_2 \quad \text{و} \quad \sigma^2_2 = n_2 p_2 q_2$$

فتوزيع المعاينة للفرق ما بين $(\hat{p}_2 - \hat{p}_1)$ يقترب من التوزيع الطبيعي بوسط :

$$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \quad \text{وانحراف قياسي :}$$

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \quad \text{لذا فإن القيمة المعيارية لهما :}$$

مثال (5)

إذا كانت نسبة النجاح في مادة الإحصاء في جامعة عمان الأهلية تساوي 0.8 وكانت نسبة النجاح لنفس المادة في جامعة الزيتونة تساوي 0.75 ، سحبت عينة عشوائية حجمها 70 طالبا من جامعة عمان الأهلية وعينة ثانية من جامعة الزيتونة حجمها 35 .

أوجد احتمال أن تزيد نسبة النجاح في جامعة عمان الأهلية عن نسبة النجاح في جامعة الزيتونة بمقدار 0.1 على الأكثر.

الحل

$$p(\hat{p}_1 - \hat{p}_2 \leq 0.10) \quad \text{المطلوب هو إيجاد}$$

وباستخدام النظرية أعلاه فإن هذا الاحتمال يساوي:

$$\begin{aligned} p(\hat{p}_1 - \hat{p}_2 \leq 0.10) \\ &= p \left[\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \leq \frac{0.10 - (0.8 - 0.75)}{\sqrt{\frac{(0.8) \times (0.2)}{70} + \frac{(0.75) \times (0.25)}{35}}} \right] \\ &= p(Z \leq \frac{0.05}{0.0873}) = p(Z \leq 0.573) \\ &= 0.5 + P(0 \leq z \leq 0.573) \\ &= 0.7157 \end{aligned}$$

أي أن نسبة النجاح في عمان الأهلية تزيد بـ 71% عنها عن الزيتونة.

5-3-5 توزيع المعاينة للباين Sampling Distribution for the Variance

إذا سحبنا عينات عشوائية كل منها ذات حجم n من مجتمع طبيعي تباينه σ^2 ثم أعيد الاختيار لعدة مرات و حسب تباين كل عينة S^2 فإننا نحصل على الإحصاءة S^2 .

فإن توزيع المعاينة S^2 ليس ذا مكانة عملية في الإحصاء لذا نتيجة إلى توزيع المعاينة لـ

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

هذا يعني أن توزيع المعاينة إلى التباين يخضع لتوزيع مربع كاي وبدرجة حرية $(n-1)$.

نظرية

سحبت عينة عشوائية حجمها n من توزيع طبيعي معدلته 5 وتباينه أي $X \sim N(\mu, \sigma^2)$ وكان S^2

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \quad \text{هو تباين العينة لذا فإن:}$$

هي قيمة التغير العشوائي χ^2 الذي له توزيع مربع كاي بدرجة حرية $n-1$.

الفصل الخامس

1- سحبت عينة عشوائية من جامعة عمان الأهلية وكان عدد الطلاب في كل كلية كما يلي:

600	كلية العلوم
800	كلية الهندسة
400	كلية الصيدلة
1200	كلية الآداب

فما عدد أفراد العينة من كل كلية وبنفس النسبة.

2- تخضع معدلات الذكاء لطلبة الصفوف الأساسية في عمان للتوزيع الطبيعي بوسط 107 وبتباين 64. سحبت عينة عشوائية حجمها 120 طالبا.

أوجد:

- توزيع المعاينة لهذه العينة.

- احتمال أن يقع الوسط الحسابي لمعدلات الذكاء في العينة ما بين 105 - 110.

3- تخضع أوزان الأطفال عند الولادة في بلد ما للتوزيع الطبيعي بوسط 3.3 كغم وانحراف معياري 1.1 كغم. سحبت عينة عشوائية حجمها 40 طفلا في أحد المستشفيات فما احتمال أن يزيد معدل الأوزان عن 3.6 كغم أو ينقص عن 2.8 كغم.

4- سحبت عينة عشوائية حجمها 400 طالبا في إحدى المدارس فإذا كانت نسبة النجاح هي 60%. أوجد:

- توزيع المعاينة لهذه العينة.
- احتمال أن يزيد معدل النجاح عن 65%.
- احتمال أن يقل معدل النجاح عن 50%.

5- الجدول التالي يمثل بيانات لعيتين من مجتمعين طبيعيين:

المتباين	الوسط الحسابي	الحجم	
12	70	40	العينة الأولى
16	70	20	العينة الثانية

أوجد:

- توزيع المعاينة للفرق.
- $P(\bar{X}_1 - \bar{X}_2) \geq 8$
- $P(\bar{X}_1 - \bar{X}_2) \leq 3$

6- إذا كان احتمال نجاح الطالب في مساق الاقتصاد هو 0.19 سحبت عينة عشوائية حجمها 81 طالباً. المطلوب:

- احتمال النجاح أكبر من 85%.
- $0.95 \leq \text{احتمال النجاح} \leq 0.85$

7- إذا كانت نسبة النجاح في الامتحان التوجيهي للفرع العلمي 80% وللفرع الأدبي 70%. سحبت عينة عشوائية حجمها (14460) للفرع العلمي و(16682) للفرع الأدبي. المطلوب:

- احتمال أن نسبة النجاح في الفرع العلمي أكثر من نسبة النجاح في الفرع الأدبي بمقدار 0.20.
- أن النسبتان متساوية.

8- أعطي توزيع طبيعي قياسي بوسط صفر وانحراف معياري 1 ما هو احتمال:

$$1- \text{ أن } Z \text{ أقل من } 1.5$$

$$2- 1.57 \leq Z \leq 1.84$$

$$3- \text{ أن } Z \text{ تقع بين } 1.57 \text{ و } 1.84 .$$

6

الفصل السادس

التقدير واختبار الفرضيات Estimation and Test of Hypotheses

6-1 نظرية التقدير

6-2 اختبار الفرضيات

6-3 استخدام طريقة P-Value لاختبار الفرضيات

الفصل السادس

6

الفصل السادس التقدير واختبار الفرضيات

Estimation and Test of Hypotheses

6-1 نظرية التقدير Estimation Theory:

وصفنا الإحصاء في الفصول السابقة وبيننا أن العينة هي جزء من المجتمع وبيننا الطرق المتعددة لسحب العينة. عملية اختيار العينة أو ما يسمى Sampling كان الغاية الرئيسية من دراسة العينات هو الاستدلال منها على خواص المجتمع الذي تعود إليه هذه العينات فإن التوزيع الاحتمالي للمتغير العشوائي يعتمد على ثابت أو أكثر لا نعرف قيمها وهذه الثوابت تسمى معالم المجتمع Parameters. ففي هذا الفصل نستخلص أن الإحصاء الاستنتاجي (التقدير واختبار الفرضيات) هي من أهم المجالات لدراسة الاحتمالات وتوزيعات المعاينة وبناء على النتائج التي تم تعميمها على المجتمع واعتبرنا ما حصلنا عليه من القيم من نتائج يصلح للمجتمع. وذلك بتقدير معالم المجتمع Estimation. فيتم التقدير باختيار عينة من مجتمع وملاحظة مفرداتها ومن ثم حساب المقياس المراد لها وتعميمها على المجتمع.

للتقدير أهمية كبيرة وله مجالات تطبيقية مختلفة في مجالات متعددة منها زراعية، طبية، إنتاجية.

فهذا يعني أن التقدير غير ثابت من عينة إلى أخرى عند استخدام نفس الطريقة أما المقرر يكون ثابت إلا إذا تغيرت الطريقة ويمكن دراسة التقدير من جانبيين.

تقدير نقطي Point Estimation.

تقدير بفترة Interval Estimation.

6-1-1 التقدير النقطي Point Estimation:

يمثل التقدير النقطي تقدير معلمة المجتمع بنقطة تحسب من بيانات العينة.

فمثلاً لتقدير متوسط المجتمع والذي يرمز له بالرمز μ نستخدم وسط العينة والذي يرمز له بالرمز \bar{X} وبذلك فإن تقدير الوسط μ هو μ والذي يساوي \bar{X} .

أما تباين المجتمع والذي يرمز له بالرمز σ^2 فإن تقديره $\hat{\sigma}^2$ هو تباين العينة S^2 . وكذلك فإن تقدير النسبة للمجتمع P هو \hat{p} وهكذا.

وكما ذكرنا سابقاً فإن الوسط الحسابي والوسيط والمنوال التي تحسب من العينات ما هي إلا تقديرات للوسط الحسابي للمجتمع μ .

المعلمة Parameter: وهو مقدار ثابت ويصف المجتمع أو يصف توزيع المجتمع كالوسط الحسابي للتوزيع أو الانحراف المعياري للتوزيع.

هناك عدة طرق لتقدير المعلمة بنقطة ومنها:

- طريقة الإمكان العظمى Maximum likelihood.
- طريقة العزوم Method of Moments.
- طريقة المربعات الصغرى Method of Least Square.
- طريقة مربع "كاي" Minimum Chi-Square.

6-1-2 التقدير بفترة Interval Estimation:

تقدير فترة الثقة للوسط الحسابي للمجتمع:

إننا لا نتوقع الحصول على معالم المجتمع المقدرة بدون خطأ مهما كان التقدير علماً بأن التقدير يزداد ثقة بزيادة حجم العينة ولكن لا يوجد سبب يرد إمكانية الحصول على تقدير معالم بدون خطأ. لذا يجب إعطاء فترة معينة لتوقع وقوع معالم المجتمع داخلها، ومثل هذه الفترة تسمى فترة الثقة. ويمكن تعريفه:

تعريف: أن تحديد فترة (a, b) التي تضم معالم المجتمع Parameters باحتمال قدره $(1 - \alpha)$ والتي تسمى فترة الثقة.

$$p(a < \mu < b) = 1 - \alpha \quad \text{حيث أن:}$$

حيث أن a, b هما متغيران، a تمثل الحد الأدنى للفترة و b الحد الأعلى

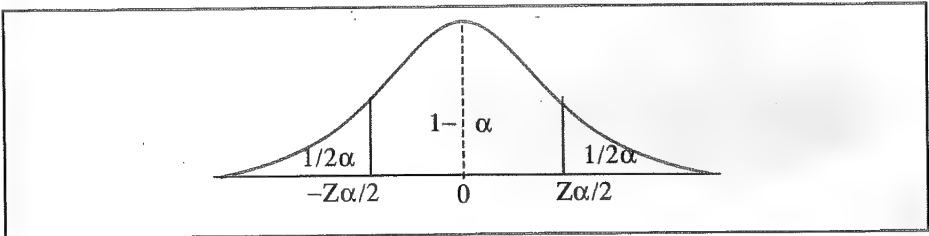
لو فرضنا أننا سحبنا عينة عشوائية حجمها n من مجتمع طبيعي $N(\mu, \sigma^2)$ وأن تباينه معلوم. وأن الوسط الحسابي للعينة هو \bar{X} والمطلوب إيجاد فترة ثقة للمعلمة μ الوسط الحسابي للمجتمع.

إن هو الوسط الحسابي للعينة الذي يخضع توزيعه إلى توزيع المجتمع التي سحبت منه لذا فإن \bar{X} يمثل المتغير العشوائي يتوزع بوسط يساوي صفر وتباين يساوي واحد. لذا فإن توزيع المعاينة هو:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

يخضع لتوزيع القياس $N(0,1)$

وأن Z هي قيمة من قيم المتغير الطبيعي القياسي Z والشكل يوضح ذلك:



احتمال أن تقع Z بين قيمتين هي $-Z \sigma/2$ و $Z\alpha/2$ بثقة $(1-\alpha)$.

$$P\left[-Z\alpha/2 < \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < Z\alpha/2\right] = 1 - \alpha$$

ويضرب الطرفين بـ $\frac{\sigma}{\sqrt{n}}$ نحصل على:

$$P\left[-Z\alpha/2 \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < Z\alpha/2 \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

وبطرح \bar{X} من كل حد ثم الضرب بسالب واحد نحصل على:

$$P\left(\bar{X} - Z\alpha/2 \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z\alpha/2 \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

ومنها نستنتج أن فترة الثقة $(1 - \alpha)$ إلى μ عندما تكون σ^2 معلومة هي:

$$\left(\bar{X} - Z\alpha/2 \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z\alpha/2 \frac{\sigma}{\sqrt{n}}\right)$$

إن هذه المعادلة تعني أن الفترة التي حدها الأيسر $\bar{X} + Z\alpha/2 \frac{\sigma}{\sqrt{n}}$ وحدها الأيمن $\bar{X} - Z\alpha/2 \frac{\sigma}{\sqrt{n}}$ والتي تحوي المعلمة μ باحتمال $(1 - \alpha)$ ومنها حدوث فترة الثقة $(1 - \alpha)$ للمعلمة μ .

مثال (1)

أجرت الاتصالات الأردنية دراسة عن فترة الثقة لاستخدام التلفون للمكالمات المحلية. لذا سحبت عينة عشوائية محلية مكونة من 15000 نداء ووجد أن معدل استخدامهم للهاتف كان يساوي 3.8 دقيقة والانحراف القياسي كان يساوي 4.0 دقيقة. قدر قيمة الوسط الحسابي للمجتمع بحدود ثقة 95%.

الحل هنا

$$\bar{X} = 3.8$$

$$\sigma = 4.0$$

$$n = 15,000$$

وباستخدام احتمال فترة الثقة أي :

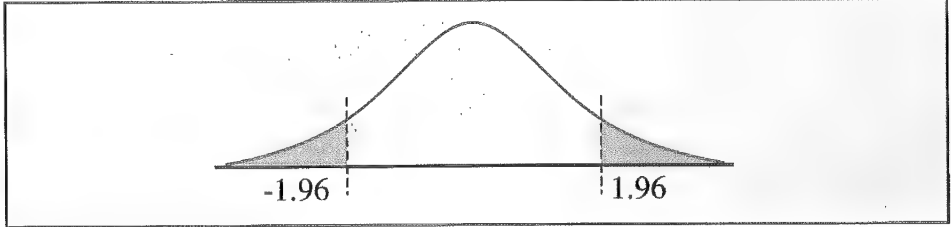
$$P\left(\bar{X} - Z\alpha/2 \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z\alpha/2 \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$1-\alpha = 95\%$$

$$\alpha = 0.05$$

$$\alpha / 2 = 0.025$$

أي أن المساحة المحصورة على طرفي توزيع الطبيعي هي 0.025 ومن هذا الاحتمال يمكن إيجاد قيم Z باستخدام جدول التوزيع الطبيعي:



$$Z_{\alpha/2} = 1.96$$

$$P\left(3.8 - 1.96\left(\frac{4}{\sqrt{15000}}\right) < \mu < 3.8 + 1.96\left(\frac{4}{\sqrt{15000}}\right)\right) = 0.95$$

$$P(3.8 - 0.06 \leq \mu \leq 3.8 + 0.06) = 0.95$$

$$3.74 \leq \mu \leq 3.86 \quad \text{وبذلك فإن}$$

ويعني أن القيمة الحقيقية للمعلمة μ تقع ما بين 3.74 و 3.86 بحدود الثقة 95%.

- تقدير فترة الثقة لـ μ عندما يكون حجم العينة كبير وتباين المجتمع غير

معلوم: Confidence Interval for Population Means (Large Sample)

ففي هذه الحالة نعوض عن الانحراف المعياري للمجتمع μ بالانحراف المعياري للعينة S أي أن:

$$P\left(\bar{X} - Z_{\alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

مثال
(2)

سحبت عينة عشوائية حجمها 50 شخصا والمتمثلة بأعمار الأشخاص المشتغلين

في سلك الجيش وكان معدل العمر 36.8 والانحراف المعياري للعمر 11.07. قدر العمر الفعلي μ لمجتمع المشتغلين بالجيش بحدود ثقة 90%.

$$1 - \alpha = 0.90$$

$$\alpha = 0.10, \alpha / 2 = 0.05$$

$$Z \alpha / 2 = Z 0.05 = 1.645$$

ومن الجدول فإن :

$$P\left(36.8 - 1.645 \frac{11.07}{\sqrt{50}} \leq \mu \leq 36.8 + 1.645 \frac{11.07}{\sqrt{50}}\right) = 0.90$$

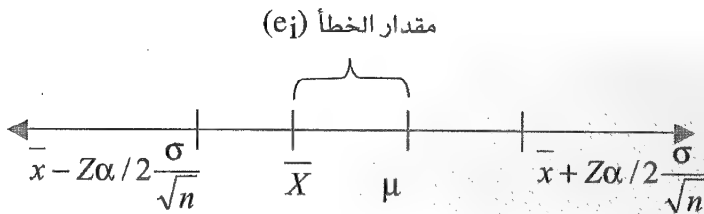
$$P(34.2 \leq \mu \leq 39.4) = 0.90$$

- تقدير حجم العينة Sample Size Consideration :

يمكن إيجاد حجم العينة في حالة إذا كانت μ تقع في منتصف الفترة معنى ذلك أن \bar{X} يعطي تقديراً لـ μ بدون خطأ (error) ولكن في معظم الأحيان فإن \bar{X} لا يكون مساوياً إلى μ بل يختلف لذلك فإن \bar{X} تختلف عن μ بكمية اقل من

$$Z \alpha / 2 \frac{\sigma}{\sqrt{n}}$$

ويمكن توضيحه كما يلي :



لذلك يمكن تقدير حجم العينة إذا كان مقدار الخطأ معلوم أي $(\bar{X} - \mu)$ والمتمثل

$$e = Z \alpha / 2 \frac{\sigma}{\sqrt{n}}$$

ومنها يمكن حساب n إذا كانت σ معلومة وقيمة e معلومة بحدود الثقة المعطاة.
وباستخدام القاعدة أعلاه يمكن إيجاد حجم العينة عندما تكون σ غير معلومة
نعوض عنها S .

مثال (3)

احسب حجم العينة التي نختارها لتكون على ثقة 95% بأن الوسط الحسابي
للمجتمع μ يختلف عن \bar{X} بأقل من 0.06 علماً بأن الانحراف القياسي للمجتمع =
0.3.

الحل

$$1-\alpha=0.95$$

$$\alpha=0.05$$

$$Z_{\alpha/2}=1.96$$

$$e=0.06$$

ومن الجدول :

$$n=\left(\frac{Z_{\alpha/2}}{e}\right)^2=\left(\frac{1.96(0.3)}{0.06}\right)^2=96$$

فإن العينة التي حجمها 96 تعطي تقدير إلى \bar{X} يختلف عن μ بأقل من 0.06.

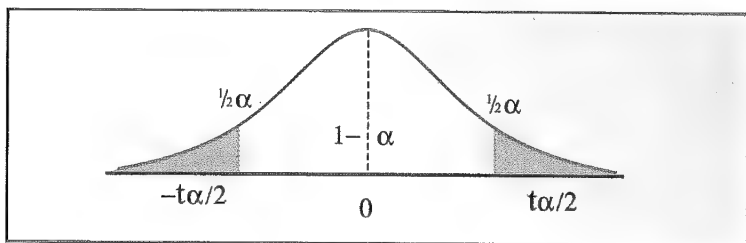
- تقدير فترة الثقة للوسط μ في حالة العينات الصغيرة:

Confidence Interval for A Population Means (Small Samples)

إن طريقة تقدير فترة الثقة للوسط الحسابي للمجتمع عندما يكون تباين
المجتمع معلوم σ^2 و حجم العينة صغيرة $n \leq 30$ فهي مشابهة لطريقة استخدام
توزيع Z في العينات الكبيرة كما ذكرنا سابقاً ولكن في هذه الحالة نستخدم توزيع t (Student's t- Distribution) كما أوضحنا في فصل تقدير المعاينة فإن التوزيع

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

فلو نظرنا إلى الشكل أدناه:



فإن احتمال T يقع ما بين $-t\alpha/2, +t\alpha/2$ وهي:

$$P(-t\alpha/2 < T < t\alpha/2) = 1 - \alpha$$

حيث أن $t\alpha/2$ هي قيمة t بدرجة حرية $v = n - 1$ وبالتعويض عن t الذي هو قريب من التوزيع الطبيعي المتغير العشوائي

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

ومنه نحصل على:

$$P\left(-t\alpha/2 < \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < t\alpha/2\right) = 1 - \alpha$$

حيث $t\alpha/2$ هي نقطة على محور اذني درجة حرية $n - 1$ التي يعلوها $\alpha/2$ في المساحة كما موضح بالشكل أي أن $t\alpha/2 = (1 - \alpha/2, n - 1)$. لذا لو ضربنا كل حد بـ $\frac{S}{\sqrt{n}}$ ثم طرحنا \bar{X} من كل حد ويضرب الناتج بـ -1 ينتج:

$$P\left(\bar{X} - t\alpha/2 \frac{S}{\sqrt{n}} < \mu < \bar{X} + t\alpha/2 \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

حيث أن الوسط الحسابي \bar{X} والانحراف المعياري S للعينة العشوائية ذات حجم $n \leq 30$ مسحوبة من مجتمع قريب من المجتمع الطبيعي.

فلذلك يمكن القول:

إذا سحبت عينة عشوائية حجمها n من مجتمع طبيعي فإن فترة ثقة $(1-\alpha)\%$ للوسط μ هي الآتي:

$$\left(\bar{X} - t_{\alpha/2} \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}} \right)$$

مثال
(4)

سحبت عينة عشوائية حجمها 9 من مجتمع طبيعي أنتجت وسطا حسابيا قدره 9.2 وبانحراف معياري 0.4، أوجد فترة ثقة 95% لمعدل المجتمع μ .

$$1-\alpha = 0.95$$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

الحل

ومن جدول توزيع t بدرجات حرية $n-1$ نجد أن

$$t_{\alpha/2} = (1-\alpha/2, n-1) = t(0.025, 8) = 2.306$$

إذن فترة الثقة 95% للمعدل μ هي:

$$P\left(\bar{X} - t_{\alpha/2} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}} \right) = 1-\alpha$$

$$P\left(9.2 - \frac{(2.306)(0.4)}{\sqrt{9}} < \mu < 9.2 + \frac{(2.306)(0.4)}{\sqrt{9}} \right)$$

$$P(8.89 < \mu < 9.51) = 0.95$$

أي أن

- تقدير فترة الثقة للفرق بين وسطين حسابيين لمجتمعين:

Confidence intervals for the difference between two means for two Population

يمكن استخدام الأسلوب السابق لإيجاد فترات الثقة للفرق بين وسطين $(\mu_1 - \mu_2)$

من مجتمعين مختلفين باستخدام نظريات توزيع المعاينة للفرق بين وسطين وفي الحالتين التاليتين:

الحالة الأولى:

- في حالة تباين المجتمعين معلومين وباستخدام النظرية التالية:

نظرية:

سحبت عينة عشوائية من مجتمع له توزيع طبيعي $N(\mu_1, \sigma_1^2)$ وعينة ثانية عشوائية من مجتمع ثاني أيضا يتوزع طبيعيا $N(\mu_2, \sigma_2^2)$ وان المجتمع الأول مستقل عن المجتمع الثاني فإن توزيع المعاينة لـ $(\bar{X}_1 - \bar{X}_2)$ يقترب من التوزيع الطبيعي بوسط حسابي قدره

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

وبانحراف قياسي قدره:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

لذا فإن :

هي قيمة المتغير الطبيعي القياسي و احتماله يقع بين القيمتين $Z\alpha/2$ و $-Z\alpha/2$

فإن فترة ثقة $(1-\alpha)$ للفرق بين الوسطين هي :

$$P \left[\bar{X}_1 - \bar{X}_2 - Z\alpha/2 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + Z\alpha/2 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

مثال

(5)

سحبت عينة عشوائية حجمها 36 من مجتمع طبيعي $N(\mu_1, 32)$ وسحبت عينة ثانية حجمها 49 من مجتمع ثاني له توزيع طبيعي $N(\mu_2, 70)$ مستقل عن الأول، فإذا كان الوسط الحسابي للعينة الأولى يساوي 47 والثانية 35، احسب:

1- فترة ثقة 59% للفرق بين متوسطيهما (μ_1, μ_2) .

2- فترة ثقة 90% إلى (μ_2, μ_1) .

	العينة الأولى	العينة الثانية
حجم العينة n	49	36
الوسط الحسابي \bar{X}	35	47
التباين σ^2	70	32

وان $1-\alpha = 0.95$

هذا يعني أن: $\alpha = 0.05$

$\alpha/2 = 0.025$

و من الجدول فإن: $Z_{\alpha/2} = 1.96$

فإن قيمة $\mu_1 - \mu_2$ باستخدام النظرية تكون كما يلي:

$$P \left[(47-35) - 1.96 \sqrt{\frac{32}{36} + \frac{70}{49}} < \mu_1 - \mu_2 < (47-35) + 1.96 \sqrt{\frac{32}{36} + \frac{70}{49}} \right] = 0.95$$

$$P(9.016 < \mu_1 - \mu_2 < 14.984) = 0.95$$

ولايجاد فترة ثقة إلى $(\mu_2 - \mu_1)$ بحدود 90% فلدينا:

$1-\alpha = 0.90$

$\alpha = 0.10$

$\alpha/2 = 0.05$

$\therefore Z_{\alpha/2} = 1.645$

$$P \left[(35-47) - 1.645 \sqrt{\frac{32}{36} + \frac{70}{49}} < \mu_2 - \mu_1 < (35-47) + 1.645 \sqrt{\frac{32}{36} + \frac{70}{49}} \right] = 0.90$$

$$P[-14.525 < \mu_2 - \mu_1 < -9.474] = 0.90$$

في حالة التباين المجتمعين (σ_1^2, σ_2^2) غير معلومة وأن حجم العينة كبيرة يمكن الاستعاضة عن تباين المجتمع بتباين العينة ليكون التقدير كما يلي:

$$P\left[\overline{X_1}-\overline{X_2}-Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}<\mu_1-\mu_2<(\overline{X_1}-\overline{X_2})+Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right]=1-\alpha$$

تطبيقات EXCEL:

يمكن إيجاد فترة ثقة للفرق بين وسطين حسابيين لعيتين مستقلين وبحجم كبير وتبايناتهم معروفة كما يلي:

EXCEL - Tools - data analysis - Z - test: Two sample for means

ملاحظة: وفي حالة عدم معرفة التباينات نعوض عنهما بتباينات العينات.

الحالة الثانية:

عندما يكون حجم العينتين صغير اصغر من 30 و أن المجتمعين مستقلين و أن σ_1^2 و σ_2^2 مجهولين فإن فترة الثقة للفرق بين وسطي مجتمعيهما يكون كما يلي:
يجب حساب تباين التجمعي للعينتين كما يلي:

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

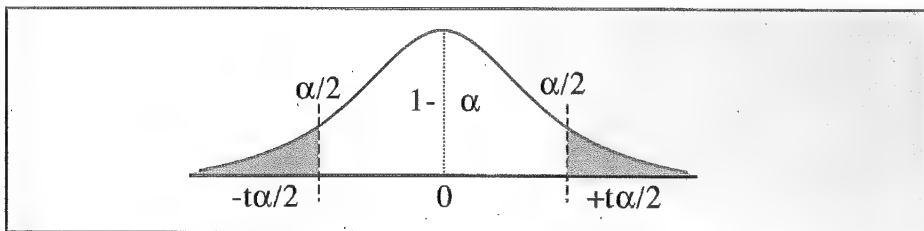
وباستخدام توزيع t وأن تكون هي النقطة المحورية بين $\alpha/2$ و $-\alpha/2$ ذي درجات الحرية $n_1 + n_2 - 2$ والتي تحصر المساحة على اليمين بـ $\alpha/2$ من المساحة أي إيجاد قيمة $t_{\alpha/2} = t(1-\alpha/2, n_1+n_2-2)$ باستخدام جدول توزيع t.

فإن فترة الثقة لـ $\mu_1 - \mu_2$ عندما تكون σ_1^2 و σ_2^2 غير معلومة وان n_1 و n_2 اصغر من 30 هي:

$$P\left[(\overline{X_1}-\overline{X_2})-t_{\alpha/2}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<\mu_1-\mu_2<(\overline{X_1}-\overline{X_2})+t_{\alpha/2}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]=1-\alpha$$

حيث \bar{X}_1 و \bar{X}_2 هما متوسط العينتين العشوائيتين المستقلتين ذات حجم n_1 و n_2 من مجتمع يقترب من التوزيع الطبيعي و S_p و هو الانحراف القياسي التجمعي وأن

$t_{\alpha/2}$ هي قيمة التوزيع t بدرجة حرية $v = n_1 + n_2 - 2$ التي تترك مساحة $\alpha/2$ إلى طرف اليمين من المنحنى كما في الشكل:



مثال (6)

سحبت عينة عشوائية من إحدى المدارس الثانوية العامة لقياس مستوى الدراسة في مادة الرياضيات ولشعبتين مختلفتين ولنفس المستوى، الشعبة الأولى عددهم 16 طالباً تم تدريسهم بالطريقة العادية والشعبة الثانية وعددهم 9 تم تدريسهم بطريقة تلفزيون وفي نهاية الفصل أعطى نفس الامتحان لكلا الشعبتين فكانت النتائج كما يلي:

	الأولى	الثانية
حجم العينة	16	9
الوسط الحسابي	85	81
التباين	4	5

أوجد فترة ثقة 95% للفرق بين متوسط الشعبتين على افتراض أن مجتمعيهما يتبع التوزيع الطبيعي وان تباينهم غير معلومة.

الحل

$$1 - \alpha = 0.95, \alpha = 0.05, \alpha/2 = 0.025$$

$$t_{\alpha/2} = (0.025, n_1 + n_2 - 2) = 2.069$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$S_p = \sqrt{\frac{15(16) + 8(25)}{16 + 9 - 2}} = 4.37$$

$$P\left[\overline{X}_1 - \overline{X}_2 - t_{\alpha/2} Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < \overline{X}_1 - \overline{X}_2 + t_{\alpha/2} Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] = 0.95$$

$$P\left[(85-81) - 2.069 \left(4.37 \sqrt{\frac{1}{16} + \frac{1}{9}}\right) < \mu_1 - \mu_2 < (85-81) + 2.069 \left(4.37 \sqrt{\frac{1}{16} + \frac{1}{9}}\right)\right] = 0.95$$

$$P = (0.2414 < \mu_1 - \mu_2 < 7.758) = 0.95$$

أي أن احتمال 95% أن الفرق الحقيقي ين متوسط الدرجات لطريقتي التعليم تقع بين 7.7-0.24.

تطبيقات EXCEL:

فترة ثقة لعينتين صغيرتين ومستقلة كما يلي:

Excel: Tools - data analysis - Test: Two Samples for means,

- تقدير فترة الثقة للفرق بين وسطين حسابيين للملاحظات المزدوجة:

Confidence Intervals for the Difference between two means for paired data.

قدير فترة ثقة للفرق بين الوسطين الحسابيين لمجتمعين للملاحظات المزدوجة أي للعينات غير المستقلة وهذه الطريقة تستخدم في حالة المقارنة بين الأزواج المتقابلة وتكون العينتين غير مستقلتين وتستخدم في حالة عندما تكون الملاحظات أزواج مرتبة ومتقابلة بحيث تكون الملاحظات مرتبطة مع بعضها البعض. مثل الأزواج (x_i, y_i) وان $i = 1, 2, \dots, n$

حيث X_i صفة لعنصر ما i و Y_i صفة أخرى للعنصر نفسه. مثلاً عند إجراء تجربة لدراسة تأثير دواء معين على ضغط الدم لعدد معين من المرضى فيقاس الضغط قبل إعطاء الدواء وبعد إعطاء الدواء. أو عند مقارنة نوعين من أصناف القمح في عدة مناطق مزرعة في لوحي متجانسين في كل منطقة وتجارب عديدة أخرى. والتجارب من هذا النوع يكون حجم العينتين متساويتين في حالة تجنب الفرق بين الملاحظات المزدوجة ونسميه d_i فإذا كان X_i صفة الملاحظة الأولى و Y_i صفة الملاحظة الثانية فإن $X_i - Y_i = d_i$ أي أن d_1, d_2, \dots, d_n هي الفرق بين n من أزواج الملاحظات فإن تقدير

نقطة $\mu_D = \mu_1 - \mu_2$ هو الوسط الحسابي للفروقات . فدرجة ثقة $1-\alpha$ لهذه الفروقات وهي μ_D يمكن إجرائها كما استخدمنا النظرية المطبقة لإيجاد فروقات الأوساط الحسابية عندما يكون حجم العينة صغير باستخدام توزيع t .

والنظرية هنا يمكن ذكرها كما يلي :

أن فروقات عينة عشوائية من مجتمع معدله μ_d وتباينه μ_d^2 فإذا افترضنا أن المجتمع يخضع إلى توزيع طبيعي $N(\mu_d, \sigma_d^2)$ من الممكن إيجاد فترة ثقة $(1-\alpha)\%$ للمعدل μ_d كما يلي:

كما ذكرنا باستخدام توزيع t فإن احتمال المتغير T يقع بين منحنى $-t\alpha/2, +t\alpha/2$ أي :

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

وأن احتمال هذا المتغير يقع بين $-t\alpha/2, +t\alpha/2$ بدرجة حرية $v = n - 1$ بالتعويض عن قيمة T

$$\therefore P\left(-t\alpha/2 < T < t\alpha/2\right) = 1 - \alpha$$

$$P\left[-t\alpha/2 < \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} < t\alpha/2\right] = 1 - \alpha$$

ويمكن تقدير الوسط الحسابي للفروقات μ_d كما سبق أي أن:

$$P\left(\bar{d} - t\alpha/2 \frac{s_d}{\sqrt{n}} < \mu_d < \bar{d} + t\alpha/2 \frac{s_d}{\sqrt{n}}\right) = 1 - \alpha$$

ويمكن حساب قيمة d و s_d كما يلي:

$$\bar{d} = \frac{\sum d_i}{n}$$

الانحراف المعياري لفروقات:

$$Sd = \sqrt{\frac{\sum d_i^2 - \frac{(\sum di)^2}{n}}{n-1}}$$

تطبيقات EXCEL:

يمكن إيجاد فترة ثقة للملاحظات المزدوجة كما يلي:

Excel: Tools - data analysis - t - test: paired two Samples for means.

مع ملاحظة استخدامه للعينات الصغيرة (أقل من 30) حيث أننا نستخدم Z-test للعينات بحجم أكبر من 30.

مثال (7)

سحبت عينة عشوائية من كلية التربية الرياضية لـ 10 طلاب وسجل عدد ضربات القلب قبل الركض وبعد الركض وكانت النتائج كما يلي:

قبل الركض X	بعد الركض Y
70	75
72	74
68	70
71	77
72	75
76	80
70	77
76	78
72	75
68	74

أوجد فترة ثقة 98% للفرق الحقيقي (هل يؤثر الركض على سرعة النبض).

$$d_i = X_i - Y_i$$

$d_i = X_i - Y_i$	d_i^2
-5	25
-2	4
-2	4
-6	36
-3	9
-4	16
-7	49
-2	4
-3	9
-6	36
-40	192

$$\bar{d} = \frac{\sum d_i}{n} = \frac{-40}{10} = -4.0$$

$$S_d^2 = \frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n-1} = \frac{192 - \frac{1600}{10}}{9} = 3.55$$

$$\therefore S_d = \sqrt{S_d^2} = 1.90$$

$$1 - \alpha = 0.98, \alpha = 0.02, \alpha/2 = 0.01$$

$$t_{\alpha/2} = 2.821$$

$$P\left(\bar{d} - t_{\alpha/2} \frac{S}{\sqrt{n}} < \mu_d < \bar{d} + t_{\alpha/2} \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

$$\left[-4 - 2.821 \frac{1.9}{\sqrt{10}} < \mu_d < -4 + 2.821 \frac{1.9}{\sqrt{10}}\right] = 1 - \alpha$$

$$P(-5.69 < \mu_d < -2.306) = 0.98$$

تقدير فترة الثقة للنسب في مجتمع Interval Estimation of Proportion:

وهو عبارة عن تقدير نقطي لنسبة النجاح في مجتمع كما ذكرنا سابقا بأن توزيع المعاينة \hat{p} (نسبة النجاحات) هو قريب من التوزيع الطبيعي بوسط حسابي قدره

$$\mu_{\hat{p}} = p \text{ وانحراف قياسي قدره } \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

وبما أن حجم العينة كبير فيمكن استخدام نظرية توزيع المعاينة للنسب أي أن:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

وان المتغير الطبيعي Z يقترب من التوزيع الطبيعي المعياري إذا كانت n كبيرة لذا

$$P(-Z\alpha/2 < Z < Z\alpha/2) \quad \text{فإن:}$$

$$\text{وبالتعويض عن } Z \text{ بما تساويها ونضرب كل حد بـ } \sqrt{\frac{pq}{n}}$$

ثم نطرح \hat{p} ونضرب الحدين بمقدار (-1) نتوصل إلى تقدير نسبة المجتمع p أي:

$$P\left[\hat{p} - Z\alpha/2 \sqrt{\frac{pq}{n}} < p < \hat{p} + Z\alpha/2 \sqrt{\frac{pq}{n}}\right] = 1 - \alpha$$

حيث أن $\hat{p} = x/n$ وتمثل عدة النجاحات في العينة التي حجمها n ومنها يمكن صياغة النظرية التالية:

نظرية:

إذا كان $\hat{p} = x/n$ نسبة النجاح في عينة عشوائية حجمها n فإن فترة الثقة $(1-\alpha)\%$ إلى p في توزيع ذي الحدين، بنسبة النجاح في المجتمع هي تقريبا:

$$\hat{p} - Z\alpha/2 \sqrt{\frac{pq}{n}} < p < \hat{p} + Z\alpha/2 \sqrt{\frac{pq}{n}}$$

حيث \hat{p} هي نسبة النجاحات في العينة العشوائية ذات حجم n و $Z_{\alpha/2}$ هي قيمة المتغير الطبيعي القياسي الذي يترك مساحة تحت المنحنى إلى اليمين بالمقدار $\alpha/2$.

مثال (8)

سحبت عينة عشوائية مكونة من 500 عائلة تملك تلفزيون في اربد و جد أن 360 منهم يملك تلفزيون ملون، احسب فترة ثقة 95% للنسبة الحقيقية للمالكي التلفزيون الملون في اربد.

الحل

$$\hat{p} = \frac{x}{n} = \frac{360}{500} = 0.72, \hat{q} = 0.28$$

$$1 - \alpha = 0.95, \alpha = 0.05, \alpha / 2 = 0.025$$

$$\therefore Z_{\alpha / 2} = 1.96$$

وباستخدام النظرية يمكن حساب قيمة P كما يلي:

$$P\left(0.72 - 1.96\sqrt{\frac{(0.72)(0.28)}{500}} < P < 0.72 + 1.96\sqrt{\frac{(0.72)(0.28)}{500}}\right) = 1 - \alpha$$

$$P(0.68 < P < 0.75) = 0.95$$

تطبيقات EXCEL:

يمكن حساب فترة ثقة لنسبة عينة واحدة كما يلي:

Excel: PH stat - one - sample test - Z - test for the proportion.

- فترة الثقة للفرق بين نسبيتين:

Confidence Interval For The difference Between Two Proportion

لإيجاد فترة ثقة إلى $p_1 - p_2$ نرجع إلى نظرية المعاينة في حالة النسب الذي نقول بأن توزيع المعاينة لـ $P_1 - P_2$ عندما n كبيرة يكون قريب من التوزيع الطبيعي بوسط

$$\mu_{p_1 - p_2} = p_1 - p_2$$

حسابي قدره:

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \quad \text{وبانحراف قياسي قدره :}$$

كما حصلنا على فترة الثقة لنسبة النجاح في مجتمع واحد. يمكن الحصول على فترة ثقة للفرق بين النسبتين شرط أن يكون حجم العينتان كبير، كما في النظرية التالية:

نظرية

إذا كانت X_1, \dots, X_n عينة عشوائية مسحوبة من مجتمع ذي الحدين $b(n_1, p_1)$ والعينة Y_1, \dots, Y_n أيضا عينة عشوائية مسحوبة من مجتمع يتوزع ذي الحدين مستقلة عن الأولى يجب أن تكون من نفس الحجم $b(n_1, p_2)$ فإن فترة الثقة $1-\alpha$ للفرق بين النسبتين $p_1 - p_2$ هي:

$$P\left(\left(\hat{p}_1 - \hat{p}_2\right) - Z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} < p_1 - p_2 < \left(\hat{p}_1 - \hat{p}_2\right) + Z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}\right) = 1 - \alpha$$

علما بأن n_1 و n_2 كبيرة وأن \hat{p}_2 هي نسبة النجاح في العينة الأولى والثانية على التوالي.

مثال (9)

سحبت عينة عشوائية حجمها 5000 رجلا من إحدى مقاطعات الولايات المتحدة (A) فوجد أن 2400 منهم يؤيد السيد الرئيس و سحبت عينة عشوائية أخرى حجمها 2000 شخصا من مقاطعة أخرى (B) وجد أن 1200 منهم يؤيد السيد الرئيس. احسب فترة ثقة 90% للفرق بين النسبة الحقيقية للمؤيدين في تلك المقاطعة.

الحل

$$\hat{p}_1 = \frac{2400}{5000} = 0.48$$

$$\hat{p}_2 = \frac{1200}{2000} = 0.60$$

$$1-\alpha = 0.90, \alpha = 0.10, \alpha/2 = 0.05$$

$$Z_{\alpha/2} = 1.645$$

وباستخدام النظرية يمكن إيجاد $p_1 - p_2$

$$\begin{aligned} & \left(\hat{p}_1 - \hat{p}_2 \right) - Z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} < p_1 - p_2 < \left(\hat{p}_1 - \hat{p}_2 \right) + Z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \\ & (-0.12) - 1.645 \sqrt{\frac{(0.48)(0.52)}{5000} + \frac{(0.6)(0.4)}{2000}} < p_1 - p_2 < (-0.12) + 1.645 \sqrt{\frac{(0.48)(0.52)}{5000} + \frac{(0.6)(0.4)}{2000}} \\ & P(-0.1414 < p_1 - p_2 < -0.0986) = 0.90 \end{aligned}$$

ومنها نستنتج أن نسبة المؤيدين في المقاطعة الثانية هي أعلى من نسبة المؤيدين في المقاطعة الأولى A.

6-2 اختبار الفرضيات Test of Hypotheses:

تصادفنا العديد من المشاكل في حياتنا اليومية ويجب اخذ القرار الملائم بشأن تلك المشاكل وبما أن اغلب الدراسات هي مستمدة من العينة المسحوبة من مجتمع. فبعد تقدير معالم هذه العينة المسحوبة من المجتمع علينا أن نعطيها ثقة اكثر، لذا نحتاج إلى اتخاذ قرار حول صحة فرضية معينة او عدم صحتها أي نحتاج إلى اختبار تلك المعالم والمتعلقة بالمجتمع. ولاتخاذ القرار الإحصائي Statistical Decision يجب النظر إلى الفروض الإحصائية Statistical Hypotheses أولاً وتوضيح بعض المفاهيم المتعلقة بها كالآتي:

الفرضية الإحصائية Statistical Hypothesis:

تعتبر بمثابة اقتراحات أولية عن معالم المجتمع غير المعلومة للباحث مستندا على البيانات المسحوبة من المجتمع عن طريق العينة Sample التي تعود إليه. فالفرضية الإحصائية هي ادعاء أو تصريح حول معلمة أو أكثر لمجتمع واحد أو لعدة مجتمعات وقد يكون هذا الادعاء صائب أو خاطئ، وعادة يتم سحب عينة من المجتمع

واستخدام المعلومات منها للوصول إلى قرار رفض او عدم رفض الفرضية الإحصائية وتقبل الفرضية في حالة أن البيانات تساند النظرية وترفض الفرضية عندما تكون بيانات العينة على خلاف ذلك.

فمثلا لمعرفة مدى تأثير إعلان معين لسلعة ما على سلوك المستهلك فيتم وضع الفرضية الإحصائية البدائية، ولتكن وجود علاقة بين تأثير الإعلان و سلوك المستهلك. هذا ويجب الملاحظة بأن قبول الفرضية الإحصائية هو ناتج عن عدم وجود أدلة كافية لرفضها من بيانات العينة ولذلك فإن قبولنا لهذه الفرضية لا يعني بالضرورة كونها صحيحة، أما إذا رفضنا الفرضية بناء على المعلومات الموجودة من بيانات العينة فإن ذلك يعني بأن الفرضية خاطئة. لذا فإن الباحث يحاول دائما أن يضع الفرضية بشكل يأمل أن يرفضها، فمثلا انه يرفض عدم وجود تأثير للإعلان على سلوك المستهلك، مثل هذه الفرضية تسمى بفرضية العدم Null Hypothesis ويرمز لها بـ H_0 . ورفضنا لفرضية العدم يقودنا إلى قبول فرضية بديلة تختلف عنها تسمى بالفرضية البديلة Alternative Hypothesis ويرمز لها بـ H_1 .

الخطأ Errors:

عند صياغة الفرضية فان طريقة اتخاذ القرار قد تؤدي إلى الوقوع في نوعين من الخطأ هما:

1- الخطأ من النوع الأول Type 1 Error: يقع الباحث في هذا النوع من الخطأ إذا تم رفض فرضية العدم عندما تكون الفرضية صحيحة ويسمى احتمال الخطأ من النوع الأول ونعبر عنه بالرمز α وان α تساوي احتمال رفض الفرضية H_0 إذا H_0 صحيحة. كما هو موضح في الجدول (1).

2- الخطأ من النوع الثاني Type 2 Error: يقع الباحث في هذا النوع من الخطأ إذا تم عدم رفض فرضية العدم عندما تكون الفرضية خاطئة ويسمى احتمال الخطأ من النوع الثاني ونعبر عنه بالرمز β أي أن β تساوي احتمال عدم رفض الفرضية H_0 إذا علم أن H_1 صحيحة كما هو موضح في الجدول.

جدول (1)

نتائج اتخاذ قرار اختبار الفرضيات

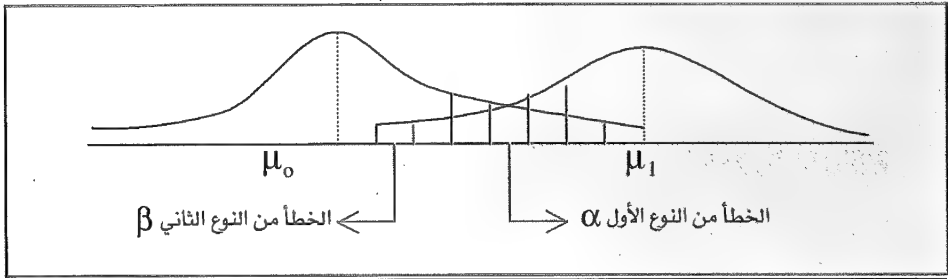
	الحالة الحقيقية	
	H_0 صحيحة	H_1 صحيحة
عدم رفض القرار H_0 رفض H_0	قرار صائب خطأ من النوع الأول	خطأ من النوع الثاني قرار صائب

مستوى المعنوية Level of Significance:

اتخاذ القرار الإحصائي بالرفض أو عدمه يتم بنسبة ما أي يعتمد على مستوى معين Level of Significant أو مستوى الاحتمال Probability Level ويرمز له α .

لذلك فمستوى المعنوية يعرف بأنه الاحتمال الذي نرفض به فرضية العدم H_0 عندما تكون صحيحة أو بعبارة أخرى هو احتمال الوقوع بخطأ من النوع الأول Type I Error ومستوى المعنوية هذه عادة تحدد من قبل الباحث و عادة تكون 0.01, 0.05, 0.1 واختيار قيمة α يعني أن نسبة الخطأ الذي صادفنا في اتخاذ القرار مثلاً 5% أي أن درجة الثقة في قرارنا تصل إلى 95% والتي تعني أن 0.95 حدود الثقة للقيمة النظرية للمجتمع (المعلمة) والقيمة الناتجة من العينة حقيقية وكبيرة.

أما قوة الاختبار Power of the Test والذي يمثل رفض H_0 عندما تكون خاطئة فيساوي $1-\beta$ ومنها نستنتج أنه كلما كانت β صغيرة زادت قوة الاختبار ويمكن القول أن هناك علاقة ما بين α و β فإذا نقص أحدهما زاد الآخر وان لحجم العينة دوراً مهماً لتقليل احتمال الخطأ لكل من α و β أي كلما زاد حجم العينة قل احتمال كلا الخطأين، علماً بأن α تحسب على أساس قيمة فرضية العدم بينما β تحسب على أساس قيمة الفرضية البديلة. ويمكن توضيح الخطأ من النوع الأول والثاني وكل من α و β بالرسم أدناه



شكل (1) يوضح الخطأ من النوع I و نوع II

خطوات اختبار الفرضيات Steps for Test of Hypothesis

سيتم توضيح خطوات اختبار الفرضيات كآلاتي:

الخطوة الأولى: تحديد توزيع المجتمع

يجب معرفة فيما إذا كان المتغير العشوائي يتوزع توزيع طبيعي، ذي حدين، أو غيره من التوزيعات فهذه نقطة مهمة لاتخاذ القرار الملائم وعلمنا بان معظم الظواهر كما ذكرنا سابقا يكون توزيعها مشابه او يقترب من التوزيع الطبيعي وخاصة إذا كانت أحجام العينات كبيرة، لذلك فإن الاستناد على الطرق المذكورة في هذا الجزء من الفصل لاختبار الفرضيات تستند على التوزيعات الطبيعية.

الخطوة الثانية: صياغة الفرضية

يتم صياغة الفرضية الصفريية H_0 والتي تمثل الفرضية المراد اختبارها والتي تعتمد تحديد قيمة لمعلمة المجتمع وتكون على الشكل التالي:

$$H_0 : \mu = \mu_0$$

حيث أن متوسط المجتمع هو μ وأن μ_0 تمثل قيمة معينة لهذا المتوسط، وعند رفض فرضية العدم يمكن اختيار فرضية بديلة H_1 بالشكل التالي:

$$H_1 : \mu \neq \mu_0$$

ويسمى الاختبار عندئذ اختبار من جهتين Two Tailed Test.

أما الاختبار الثاني للفرضية البديلة فهو $H_1 : \mu > \mu_0$

ويسمى اختبار من جهة اليمين Right-Side Test.

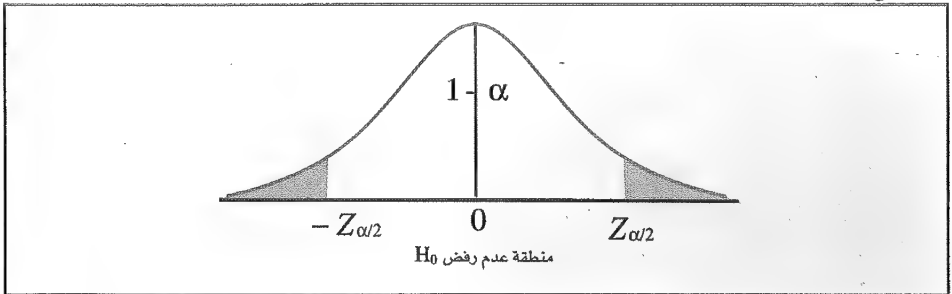
أما إذا كان الاختبار للفرضية البديلة هو $H_1 : \mu < \mu_0$

فيسمى اختبار من جهة اليسار Left-Side Test.

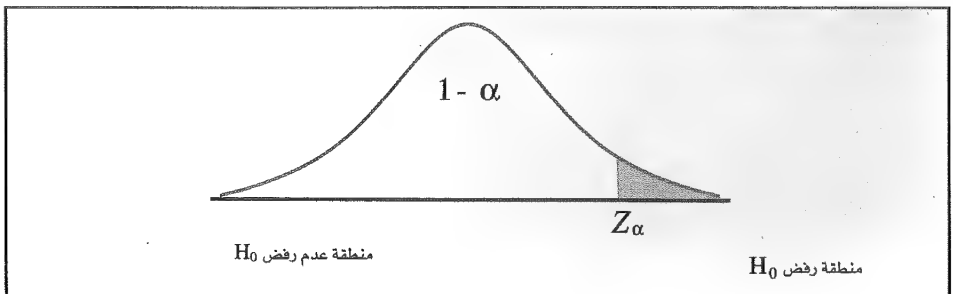
الاختبارين الآخرين من جهة اليمين أو من جهة اليسار يطلق عليهما اسم الاختبار من جهة واحدة One-Tailed Test.

الخطوة الثالثة: اختيار مستوى المعنوية α

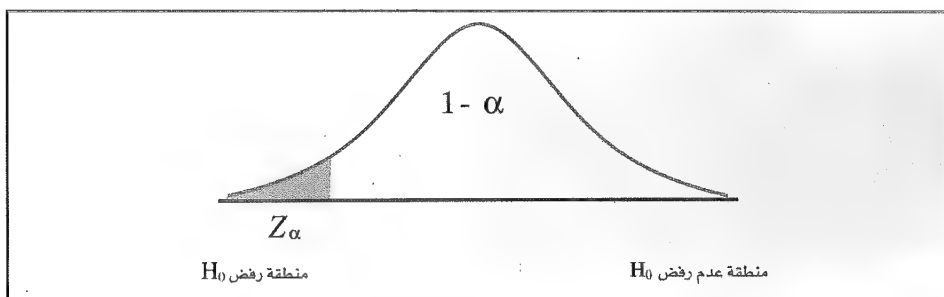
ويتم خلال هذه الخطوة تحديد قيمة α وبذلك فإن منطقة الرفض Rejection Region ومنطقة عدم الرفض Non Rejection Region ستحدد كما هو موضح في الشكل أدناه لاختبار من جهتين:



ومنطقة الرفض لاختبار من جهة اليمين سيكون بالشكل:



أما منطقة الرفض من جهة اليسار فسيكون بالشكل:



الخطوة الرابعة: إحصاء الاختبار Test Statistic

وتمثل إحصاء الاختبار قيمة محسوبة من بيانات العينة والتي على أساسها يتم الاختبار حيث يتم مقارنة التوزيع الاحتمالي المعلوم للعينة المسحوبة من المجتمع كما ذكرنا سابقا والمحدد مسبقا طبيعته مع قيمة إحصاء الاختبار المسحوبة بالتوزيع النظري (القيمة الجدولية و المسحوبة على قرار قيمة مستوى المعنوية او الدلالة α لتعيين المنطقة الحرجة او الرفض)

الخطوة الخامسة: اتخاذ القرار

أي رفض الفرضية او عدمها وهذا يتم بمقارنة إحصاء الاختبار مع منطقة الرفض فإذا وقعت ضمن منطقة الرفض نرفض الفرضية ونقبل البديلة أي أن تكون الفروق معنوية بين القيم النظرية للمجتمع و القيمة المسحوبة للعينة. وسيتم الآن اتباع الخطوات أعلاه في عملية اختبار المتوسطات وحسب الترتيب التالي:

6-2-1 اختبارات للمعلاق بالمتوسطات: حجم العينة كبير

Testing Hypotheses Concerning Mean (Large Sample)

ويتضمن هذا المبحث الاختبارات التالية:

1- اختبارات تتعلق بمتوسط واحد ويتباين المجتمع معلوم:

Testing Hypotheses for one mean (Known Variance)

الاختبار هنا يتعلق بفرضية أن الوسط الحسابي للمجتمع يساوي قيمة معينة ولتكن μ_0 أي أن:

$$H_0 : \mu = \mu_0$$

حيث أن:

μ يمثل الوسط الحسابي للمجتمع .

μ_0 هي قيمة معينة و معلومة علما بان تباين المجتمع معلوم و يساوي σ^2 .

وبالاستناد إلى أنه إذا سحبت عينة عشوائية حجمها n من مجتمع طبيعي وكانت σ^2 معلومة فان توزيع المعاينة إلى \bar{X} هو توزيع يقترب من التوزيع الطبيعي بمتوسط وتباين على التوالي :

$$\mu_{\bar{X}} = \mu , \quad \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

أما عن خطوات الاختبار فهي :

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

وهذا يعني أن الاختبار من جهتين. وبتحديد مستوى الدلالة σ المستخدم نستطيع تحديد منطقة الرفض او القبول. أما إحصاءة الاختبار تحت فرضية العدم H_0 صحيحة، فهي:

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

حيث أن Z هو المتغير الذي يخضع لتوزيع طبيعي معياري علما بان n كبيرة. وأخيرا يتم اتخاذ القرار برفض أو عدم رفض الفرضية H_0 . ويمكن تلخيص الخطوات أعلاه بشكل مخطط كما هو مبين في جدول (2) أدناه:

جدول (2)

يبين خطوات الاختبار للوسط الحسابي (باستخدام

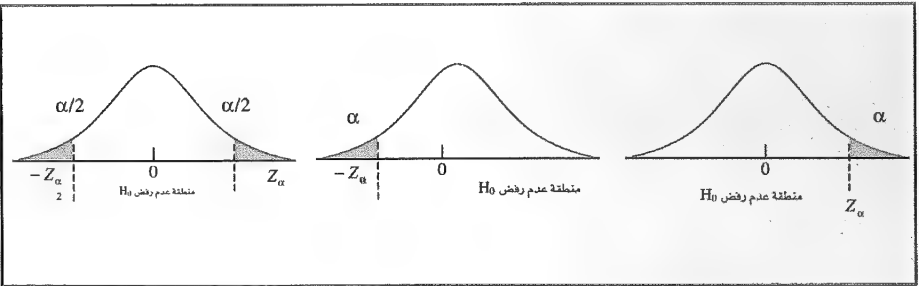
عينة كبيرة الحجم و تباين معلوم

الفرضيات (١) التوزيع الطبيعي أو حجم العينة كبير
(٢) α معلوم.

الخطوة الأولى: تحديد الفرضية $H_0: \mu = \mu_0$ الفرضية البديلة بأحد الأشكال
 $H_1: \mu > \mu_0$ $H_1: \mu < \mu_0$ $H_0: \mu \neq \mu_0$
(اختبار من جهة اليمين) أو (اختبار من جهة اليسار) أو (اختبار من جهتين)
الخطوة الثانية: تحديد مستوى المعنوية α

الخطوة الثالثة: إيجاد قيمة إحصاء الاختبار.
$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

الخطوة الرابعة: تحديد القيم الحرجة لتكون:
 Z_α أو $-Z_\alpha$ أو $\pm Z_{\alpha/2}$
(اختبار من جهة اليمين) (اختبار من جهة اليسار) (اختبار من جهتين)
وباستخدام جدول التوزيع الطبيعي والقيم الحرة نحدد مناطق الرفض لتكون:



الخطوة الخامسة: إذا كانت قيمة إحصاء الاختبار ضمن منطقة الرفض فيتم رفض H_0 ،
وبغير ذلك لا نرفض H_0 .

ملاحظة: الاختبار صحيح بالتوزيع الطبيعي وتقريبا صحيح بالعينات الكبيرة من
التوزيعات غير الطبيعية.

مثال (10)

إذا كان متوسط العلامات في إحدى المسابقات الدراسية هو 65 درجة بانحراف قياسي يساوي 10 وسحبت عينة عشوائية حجمها 40 طالبا وكان معدل علاماتهم هو 70. اختبر هل هناك فروق جوهرية بالنسبة للعلامات بمستوى دلالة 5% ؟

الحل

$$H_0 : \mu = 65$$

$$H_1 : \mu \neq 65$$

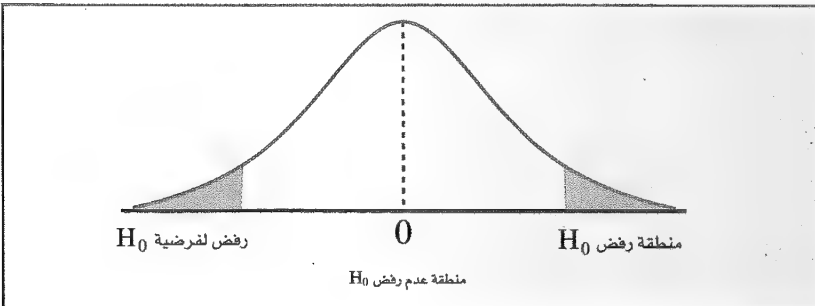
بما أن حجم العينة 40 اكبر من 30 وتباين المجتمع معلوم فان إحصاء الاختبار

هي:

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

$$Z = \frac{70 - 65}{10 / \sqrt{40}} = 3.2$$

بما أن الاختبار من جهتين فإننا نقسم α إلى منطقتين أي أن $\alpha=0.05$ يعني أن $\alpha/2=0.025$ وهذا يعني أن المساحة تحت المنحنى الطبيعي المعياري ستكون 0.025 إلى جهة اليمين من قيمة Z الموجبة و 0.025 إلى جهة اليسار من قيمة Z السالبة. وبالدخول لجدول التوزيع المعياري نجد أن $Z_{\alpha/2}=1.96$ ، وأخيرا يمكن تحديد منطقة الرفض كما في الشكل:



نقارن الآن ما بين قيمة Z من إحصاء الاختبار مع قيمة Z الجدولية أي $Z_{\alpha/2}$ ، وبما أن Z اكبر من $Z_{\alpha/2}$ لذلك فإن إحصاء الاختبار تقع في منطقة الرفض، لذا

نرفض H_0 أي أن مستوى الطلبة في العينة افضل من المستوى العام بدرجة ثقة 95%.

2- اختبار الفرضيات بمتوسط واحد لمجتمع واحد تباينه غير معلوم وحجم العينة كبير: - Testing Hypotheses for One Mean (Unknown Variance): Large Sample

في هذه الحالة بما أن حجم العينة كبير و تباين المجتمع غير معلوم فيمكن الاستعاضة عن تباين المجتمع بتباين العينة S^2 والتوزيع يكون طبيعي فيمكن استخدام المتغير Z ويكون كما يلي:

$$Z = \frac{\bar{X} - \mu_0}{S / \sqrt{n}}$$

مثال (11)

يخضع أطوال الطلبة في جامعة ما لتوزيع طبيعي وسطه 160 سم وسحبت عينة عشوائية حجمها 64 طالبا فسجلت معدلا قدره 167 سم و أنتجت انحرافا قدره 12 سم .

اختبر باستخدام $\alpha=0.05$ الفرضية:

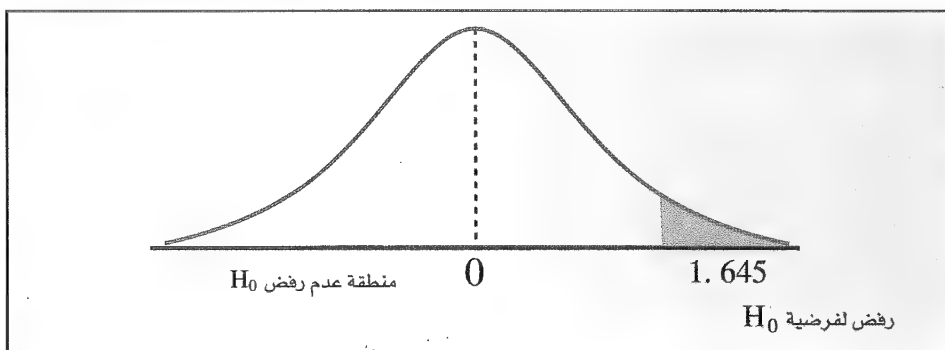
$$H_0 : \mu = 160$$

$$H_1 : \mu > 160$$

الحل

$$\begin{aligned} Z &= \frac{\bar{X} - \mu_0}{S / \sqrt{n}} \\ &= \frac{167 - 160}{12 / \sqrt{64}} = 4.67 \end{aligned}$$

وبما أن الاختبار من جهة واحدة فإن مستوى الدلالة 0.05 سيعطينا القيمة $Z_{\alpha}=1.645$ من جداول التوزيع الطبيعي. وان منطقة الرفض ستحدد بالشكل:



وأخيرا فإن القرار هو رفض H_0 أي أن المعدل أكبر من 160.

تطبيقات Excel:

يمكن اختبار الفرضيات لعينة واحدة وفي كلتا الحالتين التباين معلوم كذلك باتجاه أو اتجاهين وكما يلي:

PHstatl-one-sample test - Z test for the mean - sigma known.

3- اختبارات تتعلق بمتوسطين في حالة تباين المجتمعين معلومين: حجم العينتين كبير: Testing the Difference Between Two Means (Unknown Variances) : Large Samples

في كثير من الأحيان نرغب بأجراء مقارنات تتعلق بعيتين مسحوبتين من مجتمعين طبيعيين وبلاستناد إلى أنه إذا سحبت عينة عشوائية حجمها n_1 ومتوسطها من مجتمع طبيعي بالوسط μ_1 وتباين معلوم σ_1^2 وسحبت عينة ثانية عشوائية حجمها n_2 ومتوسطها \bar{X}_1 من مجتمع طبيعي بالوسط μ_1 وتباين معلوم σ_2^2 فإننا نختبر الفرق ما بين متوسطيهما \bar{X}_2 والمساوي إلى قيمة معينة أي:

$$H_0 : \mu_1 - \mu_2 = d_0$$

حيث d_0 هي القيمة المعينة أما الفرضية البديلة فتكون مساوية إلى واحد من الفرضيات الثلاثة التالية:

$$H_0: \mu_1 - \mu_2 \neq d_0$$

$$H_1: \mu_1 - \mu_2 > d_0$$

$$H_1: \mu_1 - \mu_2 < d_0$$

أو

أو

واختبار هذه الفروق يعتمد على توزيع المعاينة إلى $\bar{X}_1 - \bar{X}_2$ أي أن:
إحصاء الاختبار هي:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

الذي يتبع التوزيع الطبيعي بالوسط صفر و تباين واحد.
فإذا كانت $d_0=0$ فإن هذا يعني أن متوسط المجتمع للعينتين متساوي أي أن:
 $\mu_1=\mu_2$. وخطوات هذا الاختبار تلخص في جدول (3) أدناه:

جدول (3)

يبين خطوات اختبار متوسطين و تباينين معلومين: عينتين مستقلتين

الفرضيات (1) التوزيع الطبيعي او حجم العينة كبير
(2) σ_1 و σ_2 معلومين

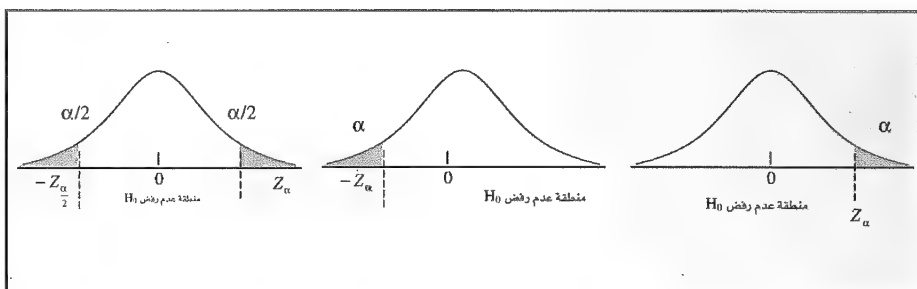
الخطوة الأولى: تحديد الفرضية	$H_0 \quad \mu_1 = \mu_2$	الفرضية البديلة بأحد الأشكال
	$H_1 \quad \mu_1 < \mu_2$	$H_1 \quad \mu_1 > \mu_2$
	$H_0 : \mu_1 \neq \mu_2$	
الخطوة الثانية: تحديد مستوى المعنوية α	(اختبار من جهة اليسار) أو (اختبار من جهة اليمين) أو	(اختبار من جهتين)

الخطوة الثالثة: إيجاد قيمة إحصاء الاختبار.
الخطوة الرابعة: تحديد القيم الحرجة لتكون:

$\pm Z_{\alpha/2}$ أو $-Z_{\alpha}$ أو Z_{α}

(اختبار من جهة اليمين) (اختبار من جهة اليسار) (اختبار من جهتين)

وباستخدام جدول التوزيع الطبيعي والقيم الحرة نحدد مناطق الرفض لتكون:



الخطوة الخامسة: مقارنة قيمة إحصاءة الاختبار مع منطقة الرفض، فإذا وقعت ضمن منطقة الرفض عندئذ نرفض H_0 ، وبغير ذلك لا نرفض H_0 .

مثال (12)

سحبت عيتان عشوائيتان مستقلتان حجمهما 50 و 35 من مجتمعين طبيعيين تباينهما $\sigma_1^2 = 136$ و $\sigma_2^2 = 136$ وكان المتوسط الحسابي للعتيتين هما:

$$\bar{X}_2 = 70 \text{ و } \bar{X}_1 = 75$$

اختبر على مستوى دلالة 5% هل توجد فروق جوهرية بين متوسط العيتين؟

الحل

$$H_0 : \mu_1 - \mu_2 = 0$$

الفرضية المراد اختبارها هي:

$$H_1 : \mu_1 - \mu_2 \neq 0$$

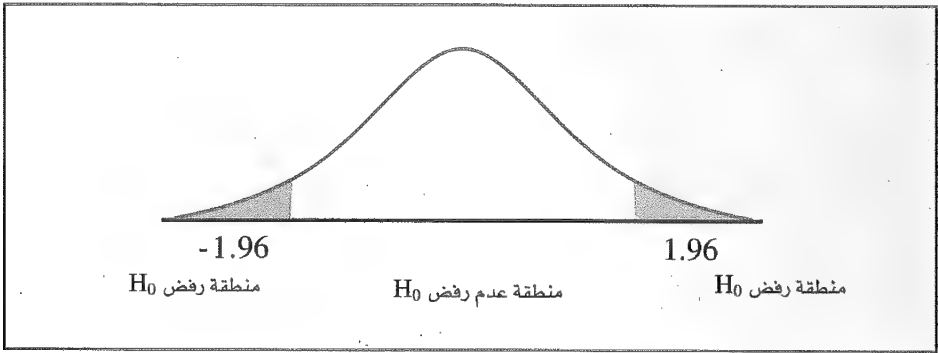
بما أن الاختبار من جهتين لذلك فإن:

$$\alpha / 2 = \frac{0.05}{2} = 0.025$$

والقيمة الجدولية $Z_{\alpha/2} = 1.96$

أما إحصاءة الاختبار فهي :

$$Z = \frac{(\bar{X}_2 - \bar{X}_1) - do}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(75 - 70) - 0}{\sqrt{\frac{136}{50} + \frac{81}{35}}} = 2.29$$



بما أن قيمة Z المحسوبة وهي 2.29 واقعة في منطقة الرفض لذا نرفض H_0 أي أن هناك فروق معنوية بين المجتمعين تحت دلالة 5%.

مثال (13)

أعطى امتحان الكفاءة في مادة اللغة الإنكليزية إلى مجموعتين من الطلبة : الأولى مؤلفة من 50 طالبة والثانية من 75 طالبا فكان متوسط درجات الطالبات هو 82 درجة بانحراف قياسي 6 بينما متوسط درجات الطلاب 70 بانحراف قياسي 7.5 .
اختبر باستخدام $\alpha = 0.01$ الفرضية:

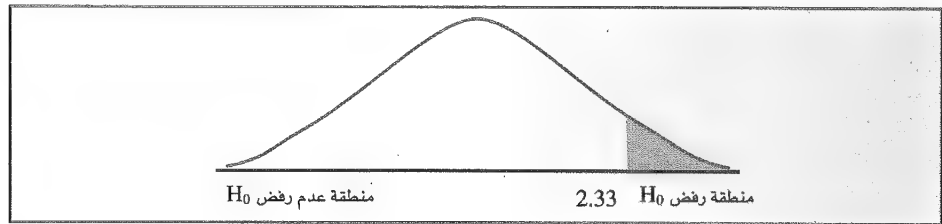
$$H_0 : \mu_1 - \mu_2 = 10$$

$$H_1 : \mu_1 - \mu_2 > 10$$

الحل

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(82 - 70) - 10}{\sqrt{\frac{36}{50} + \frac{56.25}{75}}} = 1.648$$

وبما أن الاختبار من جهة واحدة فإن $\alpha = 0.01$ يعطي $Z_{\alpha} = 2.33$ وأن منطقة الرفض تحدد بالشكل:



وبما أن قيمة Z تقع في منطقة عدم الرفض، لذلك لا نرفض H_0 وهذا يعني أن الفرق بين المتوسطين هو 10 .

4- الفرق بين متوسطين في حالة تباين المجتمعين غير معلومين بحجم

العينيتين كبير: Testing the Differences Between Two Means (Unknown Variances)-Large Sample

يمكن الاستعاضة عن تباين المجتمعين σ_1^2 و σ_2^2 بتباين العينتين S_1^2 و S_2^2 على التوالي إذا كان حجم العينتين كبيرتين. فإحصاء الاختبار تكون:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

مثال
(14)

سحبت عينة عشوائية حجمها 100 من الناجحات في الثانوية العامة من إحدى المدارس بمعدل 70 وانحراف معياري 9.5 وسحبت عينة عشوائية أخرى من مدرسة ثانية حجمها 150 طالبة وكان معدل نجاحها يساوي 65 وانحراف معياري 12. اختبر على المستوى دلالة 1% على أن:

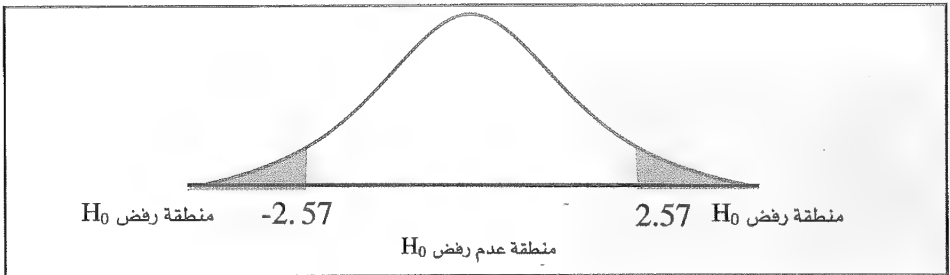
$$H_0 : \mu_1 - \mu_2 = 2$$

$$H_1 : \mu_1 - \mu_2 \neq 2$$

بما أن العينتين كبيرتين وتباين مجتمعهما غير معلوم فيمكن اخذ تباين العينتين بدلا منهما. وان إحصاء الاختبار هي:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(70 - 65) - 0}{\sqrt{\frac{(9.5)^2}{100} + \frac{(12)^2}{150}}} = 3.64$$

وبما أن الاختبار من جهتين فإن $\alpha/2=0.005$ وان قيمة Z تساوي 2.57 من جدول التوزيع الطبيعي المعياري، وبذلك فإن منطقة الرفض تحدد بالشكل:



وأخيرا فإن القرار هو رفض H_0 بما أن قيمة Z المسحوبة اكبر من القيمة الجدولية أي أن هناك فروق جوهرية ما بين الوسطين.

مثال (15)

باستخدام نفس بيانات المثال السابق اختبر باستخدام 1% الفرضية الإحصائية:

$$H_0 : \mu_1 - \mu_2 = 2$$

$$H_1 : \mu_1 - \mu_2 \neq 2$$

الحل إحصاء الاختبار هي:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(70 - 65) - 2}{\sqrt{\frac{(9.5)^2}{100} + \frac{(12)^2}{150}}} = 2.198$$

وبما أن الاختبار من جهتين فإن $\alpha/2=0.005$ وأن $Z_{\alpha/2}=2.57$ من جدول التوزيع الطبيعي المعياري فإن منطقة الرفض هي كالسابق.

وأخيرا فإن القرار هو عدم رفض H_0 بما أن قيمة Z تساوي 2.198 اصغر من قيمة Z الجدولية 2.57 أي أن الفرق يساوي 2 علامة.

6-2-2 اختبارات لتعلق بالموسطات: والبيانات غير معلومة

Testing Hypotheses Concerning Mean: Unknown Variances

ويتضمن هذا المبحث الاختبارات التالية:

1- اختبار الفرضيات لمجتمع طبيعي تباينه غير معلوم وحجم العينة صغير:

Testing Hypothesis About One Population Mean (Unknown Variance):

Small Sample size

هذه الحالة مختلفة عن اختبار الفرضيات عندما يكون حجم العينة كبيرا و التباين معلوم. فهنا حجم العينة صغير و المتفق عليه $n < 30$ وبما أن تباين المجتمع σ^2 غير معلوم نستخدم تباين العينة S^2 وان توزيع المعاينة للمتغير العشوائي يأخذ توزيع t بدلا من شكل التوزيع الطبيعي.

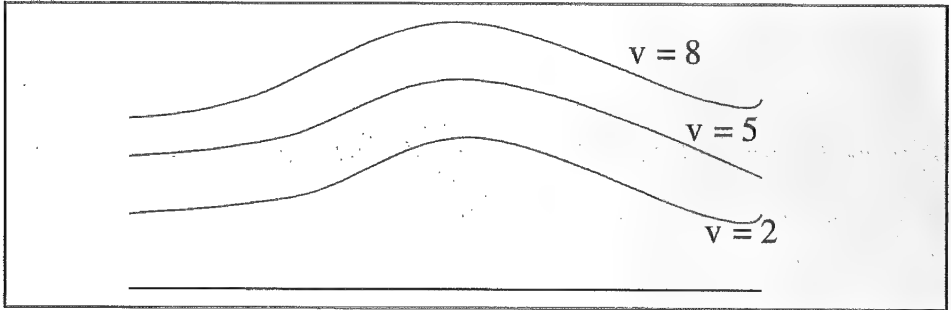
ولكن وليام كوسيت W.S.Gosset نشر بحث استطاع أن ينسق معادلة للتوزيع الاحتمالي للإحصائية T وان قيمتها هي: $t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$

خاص بالعينات الصغيرة حيث افترض Gosset بان العينات قد تم سحبها من مجتمع له توزيع طبيعي ونشره باسم مستعار Student Test أي توزيع t الذي اجري عليها التعديلات من قبل R.A. Fisher وان لهذا التوزيع له صفات مشابهة لصفات التوزيع الطبيعي. حيث أن صفات توزيع t :

- 1- إن توزيع t هو توزيع مضبوط حيث قيمة $-\infty < t < \infty$
- 2- التوزيع ذو قمة واحدة بشكل الناقوس متمائل حول الصفر.
- 3- توزيع t أكثر تفلطحاً من التوزيع الطبيعي أي أن المساحة في طرفي التوزيع أطول مما هي عليه في التوزيع الطبيعي.

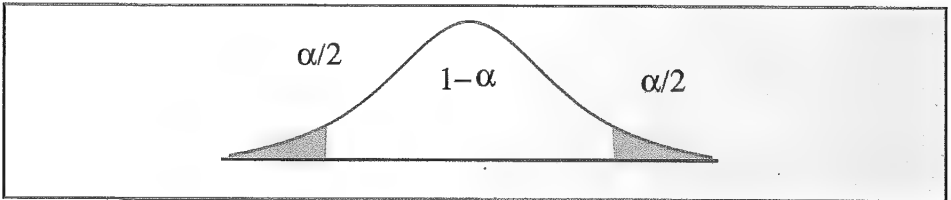
4- يعتمد على درجة الحرية (v) Degree of Freedom لا حساب احتمال.

5- إذا زاد حجم العينة فانه يقترب او يشبه التوزيع الطبيعي و الشكل التالي يبين توزيع t لدرجات حرية مختلفة، حيث أن $v = n - 1$ تمثل درجة الحرية.



أما عن كيفية استعمال جداول t فنقول:

توزيع t يعتمد على درجات الحرية $n - 1$ لتحديد المساحة الاحتمالية له و كذلك يجب معرفة مستوى المعنوية σ . فإن جدول t يتألف من العمود الأول و المتمثل بـ v درجة الحرية أما الصف فيمثل قيمة مستوى الدلالة او المعنوية σ أما الأرقام التي تحت الأعمدة فتمثل المساحة ويكون الجدول ذو طرفين كما في الشكل أدناه:



فمثلا باستخدام عينة عشوائية بالحجم 10 و باستخدام مستوى معنوية $0.05/2$ باستخدام جدول توزيع t من الملحق نجد قيمة $t(0.05, 2, 9)$.

2- اختبارات تتعلق بمتوسط واحد وتباين غير معلوم وحجم العينة:

Testing the Hypotheses About One Mean (Unknown Variance) :

Small Sample

الفرضية تتضمن هنا مقارنة متوسط المجتمع μ مع قيمة معينة μ_0 عندما يكون تباين المجتمع σ^2 غير معلوم و حجم العينة صغير $n \leq 30$ وصياغة الفرضية كما تم

بالسابق ولكن هذا الاختبار يعتمد على توزيع المعاينة لـ \bar{X} الذي يقترب من التوزيع الطبيعي حيث أن:

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

وإن t يقترب من توزيع t بدرجة حرية $v = n - 1$.

لذا فإن خطوات اختبار الفرضية حول μ هي كما تم ذكره سابقا من تحديد للفرضية الإحصائية ومستوى المعنوية وباستخدام إحصاء الاختبار ومقارنتها بمنطقة الرفض يتم اتخاذ القرار.

مثال (16)

إذا علمنا أن معدل تحصيل الطلبة في مقرر الإحصاء هو 50. اختبر الفرضية مقابل فرضية أن معدل الطلبة يختلف عن 50 باستخدام عينة عشوائية مسحوبة من مجتمع طبيعي إلى 15 طالب وكان معدل درجاتهم 53 بانحراف معياري 7 ومستوى دلالة 5%.

الحل

أولا يتم تحديد الفرضية المراد اختبارها H_0 والفرضية البديلة كآلاتي:

$$H_0: \mu = 50$$

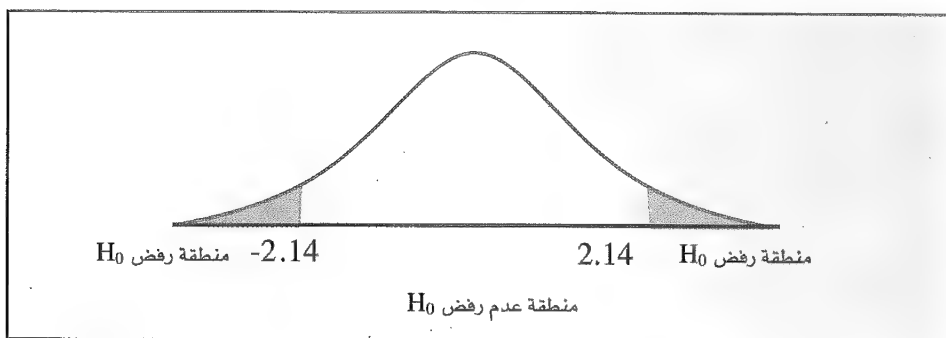
$$H_1: \mu \neq 50$$

أما إحصاء الاختبار فهي :

$$t = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} = \frac{53 - 50}{7 / \sqrt{15}} = 1.67$$

وبما أن الاختبار من جهتين فإن $\alpha/2 = 0.005$ وإن $t_{(0.025, 14)} = 2.14$ و $t_{(\alpha/2, v)}$

أخيرا فإن منطقة الرفض تحدد بالشكل :



و بما أن قيمة t المحسوبة هي اصغر من القيمة الجدولية أي تقع في منطقة عدم الرفض، لذلك لا نرفض الفرضية القائلة أن معدل الطلبة هو 50 .

لتطبيقات Excel:

يمكن اختيار الفرضيات لعينة والتباين غير معلوم أما الحجم فلا يهم إن كان كبيراً أم صغيراً كما يلي:

Excel - PHstat - one - sample - tests - t - test for the mean - sigma known

3- اختبار الفرضيات المتعلقة بالفرق ما بين متوسطين:

Testing for the Differences Between Two Means

عندما يكون حجم العينة صغير وتباين المجتمعين σ_1^2 و σ_2^2 غير معلومين نستخدم اختبار t وبالأستناد إلى أنه إذا كانت \bar{X}_1 و S_1^2 هما الوسط الحسابي والتباين للعينة العشوائية الأولى ذات حجم n_1 والمسحوبة من مجتمع يتوزع طبيعياً بمتوسط μ_1 وتباين غير معلوم σ_1^2 . وان \bar{X}_2 و S_2^2 هما الوسط والتباين للعينة العشوائية الثانية ذات حجم n_2 والمسحوبة من مجتمع طبيعي له متوسط μ_2 وتباين غير معلوم σ_2^2 فان توزيع t هو:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

حيث أن t تتوزع توزيع t بدرجة حرية $v = n_1 + n_2 - 2$ علماً أن Sp يمثل التباين التجميعي للعينتين العشوائيتين المستقلتين وأن:

مثال (17)

من سجلات مستشفى ابن الهيثم للولادة سحبت عينة عشوائية مؤلفة من 21 طفلا ذكرا حديثي الولادة

فكان معدل أوزانهم 3.1 كغم بانحراف معياري 1.7 وسحبت عينة أخرى مكونة من 18 طفلة من الإناث حديثي الولادة فكان معدل أوزانهم 2.9 بانحراف معياري 1.9 كغم. علما بأن كلا الوزنين يخضعان إلى التوزيع الطبيعي وأن تباين مجتمعيهما غير معلوم اختبر:

- a. $H_0 : \mu_1 = \mu_2$
 $H_1 : \mu_1 > \mu_2$
 $\alpha = 0.05$
- b. $H_0 : \mu_1 - \mu_2 = 0.3$
 $H_1 : \mu_1 - \mu_2 \neq 0.3$
 $\alpha : 0.01$

الحل

لاختبار الفرضية الأولى وهي:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 > 0$$

وبما أن الاختبار من جهة واحدة فإنه وباستخدام جدول t نجد أن:

$$t_{(0.05, n_1 + n_2 - 2)} = t_{(0.05, 28)} = 1.701$$

ولاستخرج التباين التجمعي للعينتين نجد:

$$S^2_p = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(12 - 1)(1.7)^2 + (18 - 1)(1.9)^2}{12 + 18 - 2} = 3.327 \quad \text{وأن:}$$

$$Sp = \sqrt{3.327} = 1.82$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{أما إحصاء الاختبار فهي:}$$

$$= \frac{(3.1 - 2.9) - 0}{1.82 \sqrt{\frac{1}{12} + \frac{1}{18}}} = 0.3$$

وأخيرا فإن القرار عدم رفض H_0 بما أن قيمة t المحسوبة 0.3 اصغر من قيمة t الجدولية 1.701 أي لا توجد فروق جوهرية بين الأوزان.

أما اختبار الفرضية الإحصائية الثانية وهي:

$$H_0 : \mu_1 - \mu_2 = 0.3$$

$$H_1 : \mu_1 - \mu_2 \neq 0.3$$

وبما أن الاختبار من جهتين فإن $\alpha/2 = 0.005$ وان $t_{(\alpha/2, n_1+n_2-2)} = 2.763$ وباستخدام جدول توزيع t . أما إحصاء الاختبار فهي:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{(3.1 - 2.9) - (0.3)}{1.82 \sqrt{\frac{1}{12} + \frac{1}{18}}} = \frac{-0.1}{0.67} = -0.149$$

بما أن قيمة t المحسوبة تقع ضمن منطقة عدم الرفض لذلك لا نرفض H_0 .

4- اختبارات تتعلق بمتوسطين العينتين غير مستقلتين (المشاهدات

المزدوجة): Testing the Differences Between Two Population Means:

Paired Difference Experiment

المشاهدات المزدوجة تكون على شكل أزواج بحيث أن الاختلافات الموجودة ضمن الأزواج اقل مما هي بين الأزواج وما دامت المشاهدات على هيئة أزواج فان حجم العينتين يكون متساوي وصغير، وهناك تطبيقات كثيرة مترتبة على هيئة أزواج متقابلة مثل (X_i, Y_i) و $i = 1, \dots, n$ حيث X_i تمثل صفة للعنصر i و Y_i صفة ثانية للعنصر نفسه فمثلا التجارب المتعلقة بمرضى السكري فإن البيانات هي قياس نسبة السكر في الدم قبل وبعد إعطاء الدواء. او مقارنة أسلوبيين مختلفين من طرق التدريس... الخ من التجارب، ويمكن توضيحها كما في الجدول التالي:

الأزواج	X_i	Y_i	$d_i = x_i - y_i$
1	x_1	y_1	$x_1 - y_1$
2	x_2	y_2	$x_2 - y_2$
\vdots	\vdots	\vdots	\vdots
n	x_n	y_n	$x_n - y_n$

فإذا كانت X_i عينة عشوائية مسحوبة من مجتمع بمعدل μ_x و Y_i عينة عشوائية ثانية من مجتمع بمعدل μ_y فان اختبار الفرق ما بين المجتمعين $\mu_x - \mu_y$ هو:

$$H_0: \mu_x - \mu_y$$

بشرط أن العينتين غير مستقلتين فإن خطوات العمل هي:

الخطوة الأولى نحسب معدل الفروق ما بين العينتين:

$$d_i = x_i - y_i$$

$$\bar{d} = \frac{\sum d_i}{n}$$

الخطوة الثانية نحسب تباين الفروق

$$S^2 d = \frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n-1}$$

الخطوة الثالثة: نستخدم إحصاء الاختبار t بالشكل

$$t = \frac{\bar{X} - \mu_d}{\frac{Sd}{\sqrt{n}}}$$

حيث أن d_0 هي القيمة التي يمكن أن يأخذها μ_d وأن t هي قيمة من التوزيع الطبيعي بدرجة حرية $v = n - 1$ ونحسب منطقة الرفض حسب نوع الفرضية البديلة H_1 كما ذكرنا سابقا.

$H_1 : \mu_d \neq d_0$ فإذا كانت:

$|t| \geq t(\alpha/2, v)$ إن :

$H_1: \mu_d > d_0$ وإذا كانت:

$t > t(\alpha/2, v)$ فإن :

$H_1: \mu_d > d_0$ أما إذا كانت:

$t < -t(\alpha, v)$ فإن :

ويلاحظ بأن هذا الاختبار هو مشابه تماما للاختبار باستخدام عينة واحدة.

مثال (18)

أراد باحث مقارنة الرواتب التي يستلمها الخريجين الجدد للأزواج والزوجات وكانت النتائج كما يلي:

الأزواج	1	2	3	4	5	6	7	8	9	10
الزوج	24.3	26.5	25.4	23.5	28.5	22.8	24.5	26.2	23.4	24.2
الزوجة	23.8	26.6	24.8	23.5	27.6	23.0	24.2	25.1	23.2	23.5

اختبر باستخدام $\alpha/2=0.05$ الفرضية :

$$H_0 : \mu d = 0$$

$$H_1 : \mu d > 0$$

الحل

علينا أولاً استخراج d_i و d_i^2 كما في الجدول التالي:

	d_i	d_i^2
1	0.5	0.25
2	-0.1	0.01
3	0.6	0.36
4	0	0.0
5	0.9	0.81
6	-0.2	0.04
7	0.3	0.09
8	1.1	1.21
9	0.2	0.04
10	0.7	0.49
	4.00	3.300

ومن ثم نستخرج قيمة S^2d و Sd على التوالي بالشكل

$$\bar{d} = \frac{\sum d_i}{n} = \frac{4.000}{10} = .400$$

$$S^2 d = \frac{3.3 - (0.4)^2 / 10}{10 - 1} = 0.36$$

$$Sd = 0.6$$

أما إحصاء الاختبار فهي

$$t = \frac{\bar{X} - \mu d}{Sd / \sqrt{n}} = \frac{0.4 - 0}{0.6 / \sqrt{10}} = 2.11$$

وبما أن الاختبار من جهة واحدة وباستخدام توزيع t نجد أن

$$t(\alpha, v) = t(0.05, 9) = 1.83$$

وبما أن قيمة t المحسوبة 2.11 اكبر من الجدولية 1.83 لذا نرفض H_0 أي أن متوسطيهما غير متساوي.

6-2-3 اختبارات للمقارنات بالنسب:

Testing Hypothesis Concerning Proportion

ويتضمن هذا المبحث نوعين من الاختبارات كآلاتي:

1- اختبارات حول نسبة واحدة من توزيع ذي حدين: Testing Hypothesis for

One Proportion :Binomial Distribution

إذا كان X متغير عشوائي يتبع توزيع ذي الحدين Binomial Distribution وأن الفرضية الإحصائية هنا متعلقة بالنسبة p لمجتمع ذي خاصية معينة فإن هذا الاختبار يشبه الاختبارات المتعلقة بالوسط الحسابي ، والفرضية هنا هي مقارنة النسبة p من توزيع ذي حدين بقيمة معينة p_0 أي:

$$H_0 : p = p_0$$

حيث أن p هي معلمة توزيع ذي حدين والتي تمثل احتمال النجاح. وان p_0 هي قيمة معينة معلومة.

فإذا كان المتغير العشوائي X يتبع توزيع ذي الحدين $X \sim Bi(n, p)$ وكان حجم العينة n كبير وان p_0 لا تكون قريبة جدا من الصفر او الواحد فان توزيع X يقترب من التوزيع الطبيعي وبذلك فان إحصاء الاختبار هذا التوزيع تكون كما يلي:

$$Z = \frac{X - np_0}{\sqrt{np_0q_0}} = \frac{p - p_0}{\sqrt{p_0q_0 / n}}$$

$$\mu = np_0, \sigma = \sqrt{np_0q_0}$$

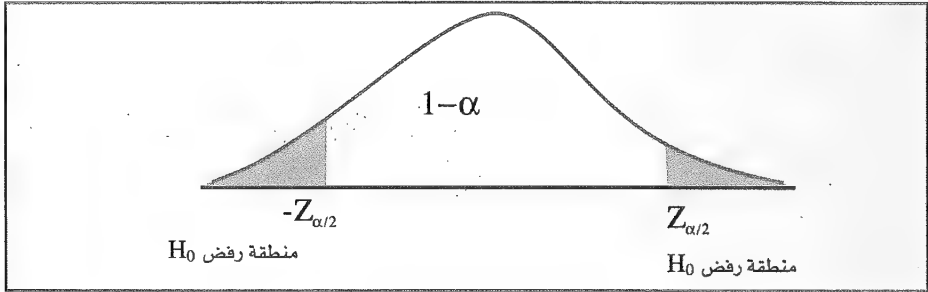
فإذا كانت الفرضية H_0 صحيحة فإن توزيع المعاينة إلى Z هو قريب من التوزيع

الطبيعي المعياري تحت مستوى دلالة معينة تساوي α إذا كانت الفرضية البديلة هي:

$$H_1 : p \neq p_0$$

فإن منطقة الرفض هي $Z > \alpha/2$ أو $Z < -Z_{\alpha/2}$.

أما شكل التوزيع مع منطقة الرفض فموضح أدناه:

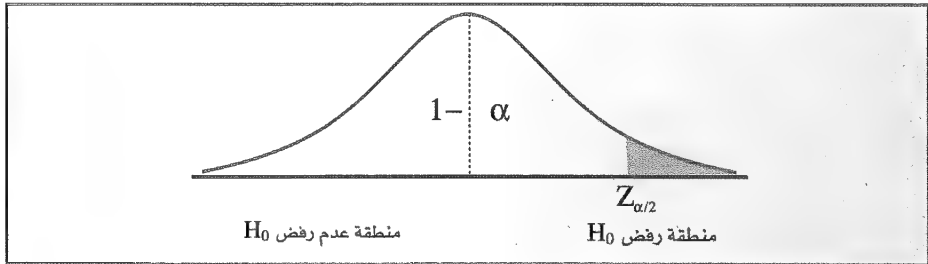


أما إذا كانت الفرضية البديلة:

$$H_1 : p > p_0$$

فإن منطقة الرفض تكون $Z > Z_{\alpha}$

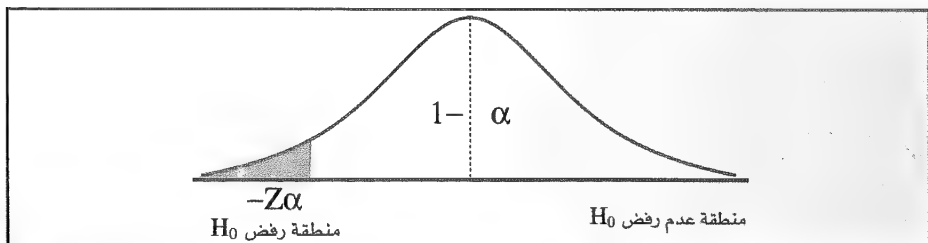
وشكل التوزيع هو:



أما إذا كانت الفرضية البديلة هي:

$$H_1 : p > p_0$$

فإن منطقة الرفض تكون $Z < -Z_{\alpha}$ وشكل التوزيع يكون:



مثال (19)

إذا كانت نسبة مستخدمي النظارات الطبية في المدارس الثانوية تساوي 60%. سحبت عينة عشوائية حجمها 100 طالبة فوجد أن 50 منهم يستخدم النظارات الطبية. اختبر على مستوى 5% أن:

$$H_0 : p = 0.60$$

$$H_1 : p \neq 0.60$$

الحل

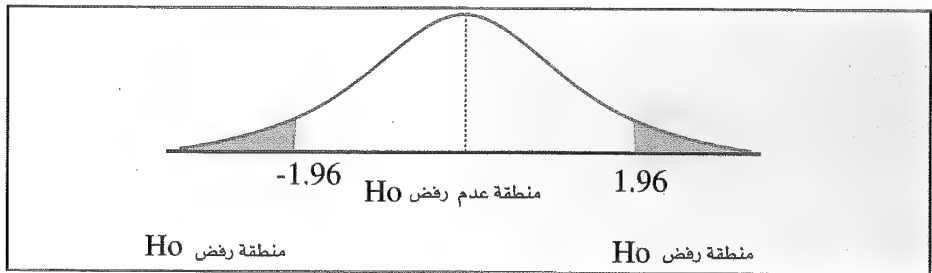
نجد أولاً قيمة \hat{P} وتمثل:

$$\hat{P} = \frac{X}{n} = \frac{50}{100} = 0.50$$

وباستخدام $p_0 = 0.60$ و $q_0 = 0.40$ نجد إحصاء الاختبار لتكون:

$$Z = \frac{0.50 - 0.60}{\sqrt{(0.6)(0.4)/100}} = \frac{-0.10}{0.048} = -2.08$$

وبما أن الاختبار من جهتين فإن $Z_{\alpha/2} = 1.96$ من جدول التوزيع الطبيعي. وبما أن قيمة Z واقعة في منطقة الرفض، لذلك نرفض H_0 حيث أن:



تطبيقات EXCEL:

يمكن اختيار الفرضية للنسبة لعينة واحدة فقط كما يلي:

PHstat - one sample tests - Z test for the proportion.

2- اختبارات تتعلق بالفرق بين نسبتي: Testing Hypothesis for the

Difference Between Two Proportion

الفرضية هنا تشمل الفرق بين معلمتين (نسبتين) لتوزيع ذي الحدين فإذا كان حجم العينتين كبير فيمكن القول أنه إذا سحبت عينة عشوائية حجمها n_1 من مجتمع يتوزع ذي الحدين $bi(n_1, p_1)$ وسحبت عينة عشوائية ثانية حجمها n_2 من مجتمع يتوزع ذي الحدين $bi(n_2, p_2)$ فإن الفرضية المراد اختبارها هي:

$$H_0 : p_1 = p_2$$

$$H_0 : p_1 - p_2 = 0$$

$$H_0 : p_1 - p_2 = d_0$$

أي أن

أو

وبذلك فإن d_0 أما أن تكون مساوية إلى الصفر أو أكبر من الصفر . وسوف نناقش

الحالتين،

الحالة الأولى عندما:

$$H_0 : p_1 = p_2$$

$$H_0 : p_1 - p_2 = d_0$$

أي أن

فإن إحصاء الاختبار هي:

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - d_0}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

وأن Z هي قيمة من قيم المتغير الطبيعي القياسي عندما تكون الفرضية H_0 صحيحة وأن n_1 و n_2 كبيرة لذا ولايجاد قيمة P فيجب دمج العيتين مع بعضهما أي أن:

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

حيث أن x_1 تمثل حالات النجاح في العينة الأولى .
وان x_2 تمثل حالات النجاح في العينة الثانية.

$$\hat{q} = 1 - \hat{p}$$

أما منطقة الرفض فتعتمد على الفرضية البديلة وبمستوى الدلالة α والتي يمكن أن تأخذ أحد الحالات الثلاثة المذكورة مسبقا وهي :

$$H_0 : p_1 \neq p_2$$

$$H_0 : p_1 > p_2$$

$$H_0 : p_1 < p_2$$

أو

أو

مثال (20)

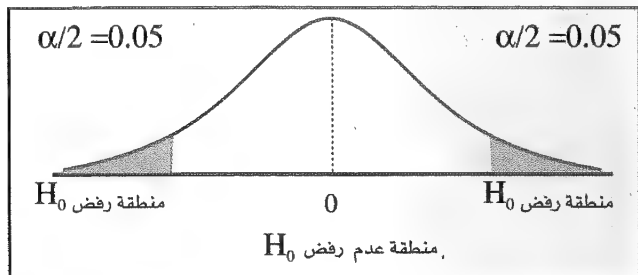
إحدى شركات تصليح السيارات رغبت في معرفة فروقات النسب لتصلح موديلين من السيارات سحبت عينة عشوائية حجمها 400 شخصا من مالكي سيارات الموديل (I) و وجدت أن 53 منهم قدم شكوى للتصليح وسحبت عينة عشوائية ثانية حجمها 500 شخصا من مالكي سيارات الموديل (II) و وجد أن 78 منهم قدم شكوى للتصليح. اختبر على مستوى 10% أنه لا توجد فروق جوهرية بين نسبي العطل في الموديلين ؟

الحل

نحدد أولا الفرضية المراد اختبارها والفرضية البديلة لتكون:

$$H_0 : p_1 - p_2 = 0$$

$$H_1 : p_1 - p_2 \neq 0$$



ثم نجد إحصاء الاختبار لتكون:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\hat{p}_1 = \frac{X_1}{n_1} = \frac{53}{400} = 0.1325 \quad \text{حيث أن:}$$

$$\hat{p}_2 = \frac{X_2}{n_2} = \frac{78}{500} = 0.1650 \quad \text{وأن:}$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{53 + 78}{400 + 500} = 0.14456 \quad \text{وأخيرا فإن:}$$

$$\hat{q} = 1 - \hat{p} = 0.8544$$

وبالتعويض أعلاه نجد أن:

$$Z = \frac{(0.1325 - 0.1650) - 0}{\sqrt{(0.1456)(0.844)\left(\frac{1}{400} + \frac{1}{500}\right)}} = \frac{-0.0235}{0.0237} = -0.99$$

وبما أن قيمة Z المحسوبة تقع في منطقة القبول، لذلك نقبل H_0 .

أما في الحالة الثانية عندما تكون الفرضية:

$$H_0 : p_1 - p_2 = d_0$$

وأن $d_0 > 0$ فهذا يعني أن p_1 تختلف عن p_2 وهنا إحصاء الاختبار المناسبة للفرق

ما بين $p_1 - p_2$ العينة هو:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - d_0}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

وأن Z هو قيمة المتغير الطبيعي القياسي عندما H_0 صحيحة

مثال (21)

سحبت عينة عشوائية من القرية A مؤلفة من 200 شخص وجد أن 120 شخصا يقرأ ويكتب. وسحبت عينة عشوائية ثانية من القرية B ومؤلفة من 500 شخص ووجد أن 240 شخص يقرأ ويكتب. أراد باحث معرفة إذا كان هناك فروق في نسبة الأشخاص الذين يقرؤون للقريتين، وكان يعتقد أن نسبة الذين يقرؤون في القرية A هي أعلى من النسبة في القرية B بأكثر من 0.05

تحت المستوى $\alpha = 0.025$

الحل

علينا أولاً تحديد الفرضية المراد اختبارها والفرضية البديلة لتكون:

$$H_0 : p_1 - p_2 = 0.05$$

$$H_1 : p_1 - p_2 > 0.05$$

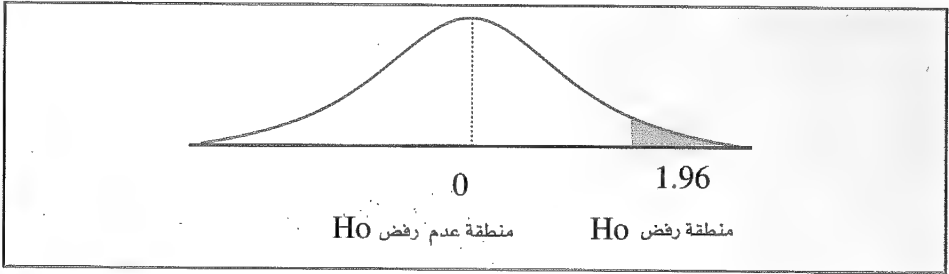
$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{120}{200} = 0.60 \quad , \quad \hat{q}_1 = 0.40 \quad \text{وأن:}$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{240}{500} = 0.48 \quad , \quad \hat{q}_2 = 0.52$$

أما إحصاء الاختبار فهي:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - d_0}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$
$$Z = \frac{(0.60 - 0.48) - 0.05}{\sqrt{\frac{(0.60)(0.40)}{200} + \frac{(0.48)(0.52)}{500}}} = 1.70$$

وبما أن الاختبار من جهة واحدة فإن قيمة $Z_{\alpha}=1.96$ باستخدام التوزيع الطبيعي وأن منطقة الرفض تحدد بالشكل:



بما أن قيمة Z واقعة في منطقة القبول كما مبين في الشكل حيث أن:

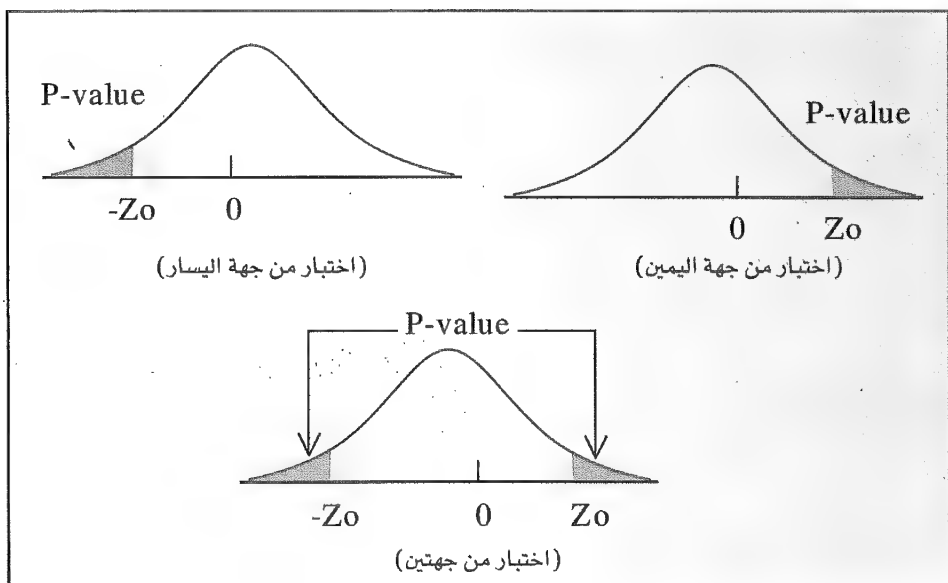
$1.70 < 1.96$ لذا لا نرفض H_0 أي أن نسبة الذين يقرؤون في قرية A لا تزيد بأكثر من 5% من نسبة من في القرية B.

6-3 استخدام طريقة P-value لاختبار الفرضيات:

Using the P-value Approach for Testing

في هذه الفقرة نتحدث عن اختبار الفرضيات باستخراج قيمة P-value والتي تمثل احتمال رفض الفرضية H_0 واستخدامها لأجل الاختبار بدلا من الطريقة المذكورة سابقا والتي كانت تعتمد على مقارنة قيمة إحصاء الاختبار (قيمة مستخرجة) مع القيم الحرجة التي تمثل مناطق الرفض (قيم جدولية). والطريقتان تعطيان نفس القرار الذي يتخذ بشأن الفرضية H_0 ولكن أسلوبهما مختلف كما نرى في الترتيب التالي:

عند حساب قيمة إحصاء الاختبار كما رأينا سابقا ولتكن Z_0 فإننا نضعها على التوزيع الطبيعي كما يظهر في الشكل آخذين بعين الاعتبار فيما إذا كان الاختبار من جهة واحدة أو جهتين.



وبعد أن يتم حساب قيمة p يتم مقارنتها مع مستويات المعنوية المناسبة فإذا كانت قيمة p صغيرة يتم رفض H_0 ، بغير ذلك لا نرفض H_0 ويتم اتباع الأسلوب نفسه لجميع أنواع الاختبارات السابق ذكرها.

أسئلة

الفصل السادس

1- عينة عشوائية من 64 مشاهدة أنتجت ما يلي:

$$\sum X = 27.4 \quad \sum X^2 = 14.3$$

$$H_0 : \mu = 0.40$$

$$H_1 : \mu = 0.40$$

اختير:

بمستوى دلالة 5%.

2- ادعت إحدى شركات إنتاج البنجر السكري بأنها أنتجت صنفا من البنجر السكري نسبة السكر فيه لا تقل عن 18% غرام لكل 100 غم وبانحراف قياس 2, 5 غم ولاختبار هذا الادعاء سحبت عينة عشوائية مؤلفة من 29 رأسا من البنجر وحسبت نسبة السكر فيها فكان وسطها الحسابي يساوي 17.2% غم لكل 100 غم فهل ادعاء الشركة صحيح عند مستوى دلالة 5%.

3- سحبت عيتين عشوائيتين من مجتمعين طبيعيين حجم الأولى 6 والثانية 5 والبيانات موضحة في الجدول التالي اختبر أن لا توجد فروق جوهرية بين متوسط مجتمعيهما بمستوى دلالة 1%.

Sample 1	Sample 2
3.1	2.3
4.4	1.4
1.2	3.7
1.7	8.9
0.7	5.5
3.4	

4- الجدول التالي يمثل بيانات عن الأجور التي تتقاضاها عيشتين عشوائيتين والمسحوبة من مجتمعين طبيعيين أحدهم يمثل المشتغلين لدى الدولة والثانية المشتغلين لدى القطاع الخاص.

الأجور لدى القطاع الخاص	الأجور لدى قطاع الدولة	حجم العينة
35	30	
35,558.97\$	33,335.20\$	الوسط الحسابي
14940.88\$	15129.09\$	الانحراف المعياري

اختبر على مستوى دلالة 5% هل توجد فروق جوهرية بين مجتمعيهما.

5- إحدى شركات المنتجات النفطية أنتجت نوع معين مطور من مادة الكاسولين المحسن لزيادة عدد الكيلو مترات من المسافات المقطوعة. ولاختبار هذا، سحبت عينة عشوائية من 10 سيارات وسارت باستخدام الكاسولين المحسن والعادي والجدول التالي يبين المسافات المقطوعة من قبل العشر سيارات. اختبر على مستوى 0.05 انه لا توجد فروق جوهرية وإن كانت هناك فروق قدرها بحدود الثقة السابقة في السؤال.

كاسولين	
محصن	عادي
25.7	24.9
20.5	18.8
28.4	27.7
3.7	13.0
18.8	17.8
12.5	11.3
28.4	27.6
8.1	8.2
23.1	23.1
10.4	9.9

6- لمعرفة فيما إذا كان هناك فروق في نسبة أصوات الناخبين لمدينين في أمريكا B و A حول ترشيح السيد بوش للرئاسة سحبت عينة عشوائية حجمها 200 شخص من المدينة A فكان 120 منهم يؤيد السيد بوش وسحبت عينة أخرى حجمها 500 شخص من المدينة B فوجد أن 240 منهم يؤيد السيد بوش. فهل تعتقد أن نسبة الأصوات التي سيحصل عليها السيد بوش من المدينة A أعلى من نسبة الأصوات في المدينة B على مستوى دلالة 5%.

7- الجدول التالي يمثل بيانات عن الأجور التي تتقاضاها عيتين عشوائيتين والمسحوبة من مجتمعين طبيعيين أحدهم المشتغلين لدى الدولة والثانية لدى القطاع الخاص.

الأجور لدى القطاع الخاص	الأجور لدى القطاع الدولة	حجم العينة
35	30	حجم العينة
35,558.97\$	33,335.20\$	الوسط الحسابي
14940.88\$	15129.09\$	الانحراف المعياري

احسب فترة ثقة 59% للفرق بين متوسطي مجتمعهما.

8- في استفتاء اجراه أحد الباحثين بشأن موضوع الاستنساخ سحبت عينة عشوائية 100 شخص فكان 40 منهم موافق. فهل يتفق ذلك مع الفرضية القائلة إن نسبة المعارضين لا تختلف عن نسبة المؤيدين وذلك عند مستوى معنوية $\alpha=0.01$.

9- اجري اختبار في إحدى المسابقات ومع شعبتين مختلفتين الأولى مكونة من 40 طالبا والثانية 30 طالبا وجد أن متوسط علامات الطلبة في الشعبة الأولى 65 درجة بانحراف معياري قدره 10 درجة ومتوسط علامات الطلبة في الشعبة الثانية 57 درجة بانحراف معياري 6 درجة اختبر هل هناك فروق جوهرية بين مستوى الطلبة في الشعبتين على مستوى معنوية $\alpha=0.05$.

10- لمعرفة إذا كان هناك فروق جوهرية بين إنتاج مصنعين مختلفين سحبت عينة عشوائية من إنتاج المصنع الأول بحجم 200 وحدة فوجد أن 180 منها صالح للاستهلاك وكذلك سحبت عينة من المصنع الثاني حجمها 500 وحدة ووجد أن 440 منها كانت صالحة فهل تعتقد أن سبة الإنتاج في المصنع الأول تزيد عن إنتاجية المصنع الثاني بأكثر من 5% بمستوى دلالة 5%.

11- إذا كان متوسط الزيادة في وزن 12 فأرة بعد تغذيتها بطريقة معينة لمدة معينة هو 145 غم وبانحراف قياس للوسط الحسابي مقداره 2.3 غم وبمستوى احتمال 5% هل يمكن القول أن متوسط الزيادة في الوزن نتيجة التغذية على هذه الطريقة لا تقل عن 150 غم اختبر ذلك.

12- شركة توزيع المحروقات ارادت تقدير الوسط الحسابي للمجتمع بحدود ثقة 59% سحبت عينة عشوائية مكونة من 100 عائلة مستهلكة للسولار كان معدل استهلاكها ما يعادل 1103 غالون بانحراف معياري 327.8. احسب الوسط الحسابي للمجتمع الذي سحبت منه العينة ثم فسر إذا استهلكت عائلة منفردة 800 غالون هل يمكن اعتبار ذلك ممكن اختبر ذلك.

13- إحدى شركات الاتصالات أجرت بحثا حول المكالمات الطويلة فوجد أن معدل ما يدفعه المواطن للمكالمة الطويلة يساوي 17.10 دولار في الشهر. وبانحراف معياري 9.80 دولار.

- سحبت عينة عشوائية لـ 50 قائمة تلفون. أوجد احتمال أن تكون مدة المكالمة اكبر من 20 دولار.

- اختبر أن: $H_0: \mu = 21$ ضد $H_1: \mu > 21$ مستخدما $\alpha = 5\%$

14- أجريت مقارنة للأسعار في مدينتين مثل اليابان وامريكا فإذا كان سعر الفرق لبعض المواد التجارية في كل من الدولتين موضحة في الجدول أدناه:

	اليابان	امريكا
حجم العينة	$n_2 = 30$	$n_1 = 50$
معدل سعر	1224.3	1154.5
الانحراف المعياري	1843	1989

اختبر أن فروقات الأسعار هي اكبر من 002 دولار بمستوى دلالة 0, 50.

15- من المعلومات المتوفرة لديك اختبر أن $\mu_1 - \mu_2$ حيث أن $\mu_1 \neq \mu_2$ بمستوى دلالة 5%

16- احسب التقدير النفطي لعينة عشوائية حجمها 12 مسحوبة من مجتمع μ

	المجموعة الاولى	المجموعة الثانية
الوسط الحسابي	38.75	35015
الانحراف المعياري	3.2	2.7
حجم العينة	100	100

وكانت أفراد العينة هي:

X: 9 6 5 3 4 7 8 9 10 3 12 6

17- في إحدى المدارس الأساسية سحبت عينة عشوائية من الطلبة الذين سيعملون النظارات الطبية حجمها 20 طالبا فوجد أن 6 منهم يستخدم النظارات الطبية. فما تقديرك للسته الذين يستعملون النظارات الطبية في تلك المدرسة.

18- سحبت عينة عشوائية حجمها 400 مفردة من مجتمع انحرافه القياسي 30 ومعدل 160 احسب فترة ثقة 59% للوسط الحسابي للمجتمع الذي سحبت منه هذه العينة.

19- سحبت عينة عشوائية من إحدى مصانع المصابيح الكهربائية حجمها 25 مصباحا فكان الوسط الحسابي لاعمار هذه المصابيح 890 ساعة. احسب فترة ثقة 90% لمعدل أعمار المصابيح المنتجة في هذا المصنع على أن الانحراف المعياري لإنتاجية هذا المصنع هو 35 ساعة.

20- سحبت عينة عشوائية في إحدى مصانع الخيوط حجمها 60 خيطا فوجد أن معدل قوة هذه الخيوط 90.4 كغم بانحراف معياري 7.5 كغم. أوجد فترة ثقة 98% لمعدل قوة جميع الخيوط التي ينتجها ذلك المصنع.

21- سحبت عينة عشوائية مكونة من 20 طالبا من جامعة عمان الأهلية لتقدير معدل المصروف الأسبوعي لهم. فوجد أن مصروفهم الأسبوعي بالدينار الأردني كما يلي:

38	51	49	38	36	35	44	50
43	41	44	38	33	45	50	51
44	39	49	52				

- ما هو تقديرك لمعدل المصروف الأسبوعي لجميع طلبة جامعة عمان.

- احسب فترة ثقة 90% لمعدل المصروف إذا كان المصروف يخضع لتوزيع طبيعي.
22- سجلت قياسات الحموضة (PH) لعينات من ماء المطر في 10 مواقع في منطقة صناعية فكانت

3.9 3.1 5.1 3.8 4.5 3.2 4.8 3.9 4.1 3.6

احسب فترة ثقة 98% لمعدل حموضة ماء المطر لكل المناطق.

23- سجل باحث أكاديمي شمل 625 طالبا من الناجحين في شهادة الدراسة الثانوية العامة. وجد أن 35% منهم التحقوا بالجامعة و20% منهم التحقوا بكليات المجتمع ولم يكمل الباقيون دراستهم بعد الثانوية العامة.

- احسب فترة ثقة 95% لنسبة المتحقين بالكليات من المتحقين بالجامعة.
- أوجد فترة ثقة 90% للنسبة الحقيقية للطلبة الذين لم يواصلوا دراستهم بعد الثانوية.

24- سحبت عينة عشوائية حجمها 9 من توزيع طبيعي وكان $\bar{X}_1 = 13.5$ و $S_1^2 = 5$ كما العينة العشوائية التالية التي حجمها 13 مسحوبة من مجتمع له توزيع طبيعي مستقل عن الأول بوسط $\bar{X}_2 = 13.5$ و $S_2^2 = 5$

- أوجد فترة ثقة 59% للتباين σ_1^2 .

- أوجد فترة ثقة 9% للانحراف المعياري σ_2 .

- أوجد فترة ثقة 59% لـ $\frac{\sigma_2^2}{\sigma_1^2}$.

25- إذا كانت أجور مندوبي المبيعات لكل من الذكور والاناث تخضع لتوزيع طبيعي متباين 100 للذكور و 144 للاناث علما بأن المجتمعين للذكور والاناث مستقلين عن بعضهما. سحبت عينة عشوائية حجمها 12 من المندوبين (ذكور) بوسط حسابي 170 وانحراف معياري 9 وعينة عشوائية ثانية (إناث) حجمها 24 بوسط 162 وانحراف معياري 10.

- أوجد فترة ثقة 98% لكل من μ_1 و μ_2 .

- احسب فترة ثقة للفرق.

26- الجدول التالي يبين التحصيل العلمي لمجموعة من الطلبة في مدينتين مختلفتين.

الدرجة	حجم العينة	المتوسط الحسابي	الانحراف المعياري
أ	200	79	11
ب	150	73	12

احسب فترة ثقة 95% للفرق بين معدلي تحصيل الطلبة.

27- الجدول التالي يمثل الأجور التي يتقاضاها عيشتين عشوائيتين والمسحوبة من مجتمعين طبيعيين أحدهما المشتغلين لدى القطاع الخاص والثانية لدى الدولة.

العينات الأولى (الخاص)	العينات الثانية (الدولة)	حجم العينة
35	30	حجم العينة
35558.79	33335.20	معدل الأجر بالدولار
14940.88	15129.09	الانحراف المعياري

احسب فترة ثقة 95% للفرق بين متوسطي مجتمعهما.

28- إحدى شركات المنتجات النفطية

أنتجت نوع معين مطور من مادة الكاسولين المحسن والمفروض أن يزيد من عدد الكيلو مترات من المسافات المقطوعة. سحبت عينة عشوائية من 10 سيارات وسارت باستخدام الكاسولين المحسن والعادي والجدول التالي يبين المسافات المقطوعة من قبل العشر سيارات. احسب مقدار الفروقات بين المسافات المقطوعة من قبل العشر سيارات بحدود ثقة 95%.

كاسولين عادي	كاسولين محسن
24.9	25.7
18.8	20.5
27.7	28.4
13.0	3.7
17.8	18.8
11.3	12.5
27.6	28.4
8.2	8.1
23.1	23.1
9.9	10.4

29- عينة عشوائية حجمها 64 مشاهدة سحبت من مجتمع وانتجت ما يلي
 $\sum X^2 = 3566$ و $\sum X = 3566$ وقدر قيمة الوسط الحسابي لمجتمع
 هذه العينة بحدود ثقة 95%.

30- احسب حجم العينة التي تختارها لتكون على ثقة 85% وأن الوسط الحسابي
 ينحرف عن العينة بمقدار 16% علما بأن الانحراف المعياري لهذه العينة هو 0.3.

31- البيانات التالية تمثل فروقات قياس ضغط الدم قبل وبعد تناول الدواء،
 فروقات ضغط الدم = 13 - 2 - 1 - 6 4 احسب الوسط الحسابي لمجتمع
 الفروقات بمعدل ثقة 90%.

32- تم اختيار مجموعتين من الطلبة في مادة الرياضيات وسحبت عيتتين
 عشوائيتين من طلاب وطالبات إحدى المدارس وكانت النتائج كما يلي:

طالبات	طلاب	
85	18	الوسط الحسابي
4	5	الانحراف المعياري
12	10	حجم العينة

على افتراض أن المجتمعين يتوزعان قريبا من التوزيع الطبيعي وأن تباينهما غير
 معلوم ولكنهما متساويان احسب فترة ثقة 97% للفرق بين متوسط المجتمعين.

33- رغب قسم السيطرة النوعية لتقدير معدل عمر المصباح المنتج من قبل الشركة
 والمنحرف بمقدار ± 20 ساعة بحدود ثقة 95% علما بأن الانحراف المعياري كان
 100 ساعة أوجد حجم العينة.

7

الفصل السابع

تحليل التباين Analysis of Variance

7-1 مقدمة

7-2 اختبار t-test

7-3 تحليل التباين

7-4 المقارنات المتعددة

7-5 اختبار تساوي عدة تباينات

الفصل السابع

7

الفصل السابع

التقدير واختبار الفرضيات

Analysis of Variance

7-1 مقدمة Introduction

هناك حالات عديدة لأنواع مختلفة من الظواهر تحتاج الى دراسة، مثلاً لقياس فعالية عدة انواع من الاسمدة على كمية محصول معين وليكن الذرة الصفراء او على محصول الرز او على محصول القمح. والامثلة التطبيقية عديدة من تأثير المدرسة على اداء الطالب وتحصيله النهائي الى تأثير المدرس على تحصيل الطالب الى تأثير الادوية للحد من مرض معين وإعطاء هذه الانواع المختلفة من الادوية الى ايضا مجاميع مختلفة من الاشخاص وتحديد سرعة الشفاء بالايام او بالاشهر وما الى ذلك من الامثلة المختلفة. فإذا كان المطلوب هو ايجاد الفرق المعنوي بين متوسطات Means كمية المحصول نتيجة لنوعين فقط من السماد نقوم باستخدام أحد الطرق أو الاختبارات الإحصائية التي سبق ذكرها في هذا الكتاب والشائعة الاستخدام وهو اختبار T-test. اما اذا كان هناك اكثر من نوعين من السماد الى أي عدد من الأنواع أو المستويات وافرض لدينا اربعة أنواع من الأسمدة فهنا نستطيع استخدام اختبار T-test لاختبار الفرق المعنوي بين متوسطات كمية المحصول لكل زوج من الأسمدة وعندئذ نحتاج للقيام باستخدام اختبار T-test ستة مرات.

فإذا كان لدينا اربعة أنواع من الأسمدة وزعت على اربعة مجاميع من هذا المحصول وكان المطلوب استخدام مستوى المعنوية $\alpha=0.05$ لقياس الفروق المعنوية بين متوسطات المحصول للأنواع الأربعة من الأسمدة. فإن اختبارات T-test لا يمكن ان تعتبر مستقلة وذلك لاستخدام كل وسط حسابي بأكثر من اختبار وهنا توجد صعوبة بتحديد مستويات المعنوية. ولهذا فإن قرار استخدام T-test يصلح لاختبار الفرق المعنوي بين متوسطين اما عندما يكون عدد المتوسطات اكثر من اثنين فتصبح العملية

صعبة ومملة وتأخذ وقت أكثر ويزيد احتمال الخطأ بشكل أكبر، وفي هذه الحالة يكون اختبار F انسب والذي نحصل عليه من تحليل التباين ANOVA.

7-2 اختبار T-test:

مثال (1)

أراد باحث دراسة تأثير استخدام الأساليب الحديثة في التعليم على تحصيل الطلبة من الدرجات وقد تم مقارنة درجات الطلبة الذين استخدموا الأساليب الحديثة مع درجات مجموعة من الطلبة لم يستخدموا هذه الأساليب.

وكانت البيانات كالتالي:

g1 : 10 5 6 7 10 6 7 8 6 5 (مجموعة الأساليب الحديثة)

g2 : 7 3 5 7 8 4 5 6 3 2 (المجموعة التي لم تستخدم الأساليب الحديثة)

الحل

أولاً يجب ان تتوزع كل مجموعة من البيانات توزيع طبيعي.

افرض ان Y_{ij} تمثل الملاحظة i في المجموعة (العينة) j .

الان يمكن وضع الفرضيات التالية على Y_{ij} .

$$Y_{ij} = \mu_j + E_{ij} \quad i = 1, 2, \dots, n_j, \\ j = 1, 2, \dots, k.$$

الخطأ العشوائي (الأخطاء الغير مسيطر عليها : E_{ij}):

عندما تكون

$$E(\epsilon_{ij}) = 0, \quad \text{var}(\epsilon_{ij}) = \sigma^2$$

في هذه الحالة المشكلة يوجد لدينا مجموعتان (عيتان او مجتمعان) ونرغب بمقارنة متوسطاتهما μ_1, μ_2 .

أي نريد اختبار الفرضيات التالية:

$$H_0: \mu_1 = \mu_2 \text{ (فرضية العدم)}$$

$$H_0: \mu_1 \neq \mu_2 \text{ (الفرضية البديلة)}$$

وهنا نحتاج لاستخدام الاختبار الإحصائي (T-test) لاختبار هذه الفرضيات بدرجة حرية (n_1+n_2-2) وكما ذكرنا سابقا بالشكل التالي:

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\bar{Y}_1 = \frac{10+5+6+7+10+6+7+8+6+5}{10} = \frac{70}{10} = 7$$

$$\bar{Y}_2 = \frac{7+3+5+7+8+4+5+6+3+2}{10} = \frac{50}{10} = 5$$

وكذلك:

$$\sum_{i=1}^{10} Y_1^2 = 10^2 + 5^2 + \dots + 5^2 = 520$$

$$\sum_{i=1}^{10} Y_2^2 = 7^2 + 3^2 + \dots + 2^2 = 286$$

والآن نحتاج لحساب S_p^2 التباين التجميعي وهذا يتطلب حساب كلا من S_2^2 و S_1^2 التباين للمجموع وكما يلي:

$$S_1^2 = \frac{\sum_{i=1}^{n_1} Y_{i1}^2 - n_1 \bar{Y}_1^2}{n_1 - 1} = \frac{520 - 10 \times 7^2}{9} = 3.33$$

$$S_2^2 = \frac{\sum_{i=1}^{n_2} Y_{i2}^2 - n_2 \bar{Y}_2^2}{n_2 - 1} = \frac{286 - 10 \times 5^2}{10} = 4$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{30 - 36}{18} = 3.66$$

ولذلك فإن (T-test) تحسب كما يلي:

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{7 - 5}{\sqrt{3.66} \sqrt{\frac{1}{10} + \frac{1}{10}}} = \frac{2}{(1.91)(0.447)} = 2.34$$

ومن جداول توزيع T (T-distribution) توجد قيمة t الجدولية بمستوى $(\alpha=0.025)$ ولدرجة حرية $(df = 20-2 = 18)$ والتي تساوي

($T_{0.025,18} = 2.101$ الجدولية)، أي ان قيمة T المحسوبة $T = 2.34$ اكبر من قيمة الجدولية $T = 2.101$ ولهذا نرفض الفرضية التي تقول ان المتوسطان متساويان أي ان المتوسطان غير متساويان.

7-3 تحليل التباين One-Way ANOVA:

إن تحليل التباين أحادي التصنيف A one-way analysis of variance (ANOVA) يستخدم عندما يكون المتغير المستقل (Xi) independent variable هو من النوع المصنف Categorical variable والذي يمثل الطبقات او عدد المجاميع عندما تكون اكثر من مجموعتين وان المتغير المعتمد (Yij) dependent variable الذي يمثل المشاهدات يتوزع توزيع طبيعي بحيث ان البيانات هو من نوع Interval variable. والمطلوب هو اختبار الفروقات بين متوسطات المجاميع او متوسطات الطبقات وعدد الطبقات او المجاميع تحدد بالمتغير X. وهذه المجاميع او العينات من المفروض ان تكون مستقلة عن بعضها البعض. ولو فرضنا لدينا K من العينات (المجموعات) ونفرض انها مستقلة عن بعضها البعض وتخضع للتوزيعات الطبيعية ذات المعدلات $\mu_1, \mu_2, \dots, \mu_K$ وهذه العينات (المجموعات) لها نفس التباين σ^2 إضافة كونها عينات عشوائية حجم كل منها n. ونرغب في إيجاد طرق مناسبة لاختبار الفرضية التالية والتي تسمى فرضية العدم والتي تنص على عدم وجود فروقات بين المعدلات وكما يلي:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

مقابل الفرضية البديلة والتي تقول يوجد على الاقل معدل مختلف: H_1

نعبر بالرمز Y_{ij} للملاحظة ذات الرقم j المأخوذة من العينة i فتظهر المشاهدات كما في الجدول التالي:

العينات (المجموعات)	الملاحظات y_{ij}				
	1	2	3	...	n
1	Y_{11}	Y_{12}	Y_{13}	...	Y_{1n}
2	Y_{21}	Y_{22}	Y_{23}	...	Y_{2n}
3	Y_{31}	Y_{32}	Y_{33}	...	Y_{3n}
.
.
.
.
.
.
.
.
k	Y_{k1}	Y_{k2}	Y_{k3}	...	Y_{kn}

ويمكن كتابة كل مشاهدة على الشكل التالي:

$$Y_{ij} = \mu_i + E_{ij}$$

حيث ان E_{ij} تقيس انحراف الملاحظة j في العينة i عن معدل المجتمع μ_i .

$$\mu_i = \mu + \alpha_i \quad \text{وباستخدام:}$$

$$\mu = \frac{\sum_{i=1}^k \mu_i}{n}$$

حيث أن:

ولهذا يمكن كتابة النموذج أعلاه على الشكل:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \dots\dots\dots (1)$$

$$\sum_{i=1}^k \alpha_i = 0 \quad \text{تحت الشرط:}$$

حيث ان α_i تعبر عن تأثير العينة i (المجتمع i)، أما افتراضات النموذج (1) فهي: E_{ij} وتمثل الخطأ العشوائي في النموذج ولجميع المشاهدات في النموذج يجب ان تتوزع توزيع طبيعي، ومستقلة عن بعضها البعض، ومعدلها ($\mu = 0$) ولها تباين ثابت يساوي (σ^2)، وباستعمال النموذج الأخير تصبح

الفرضية المبدئية القائلة بتساوي جميع المعدلات مكافئة للفرضية التالية:

$$H_0: \alpha_1 = \alpha_2 \dots\dots\dots = \alpha_k = 0$$

مقابل الفرضية البديلة:

واحدة من α_i على الأقل لا تساوي صفر: H_1

أما اختبار هذه الفرضية فيبنى على المقارنة بين تقديرين مستقلين للتباين σ^2 ونحصل على هذين التقديرين بتقسيم التغير الكلي للبيانات إلى مركبتين حسب النظرية

$$\sum_{i=1}^k \sum_{j=1}^n (Y_{ij} - \bar{Y}_{..})^2 = n \sum_{i=1}^k (\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_{i=1}^k \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i.})^2$$

$$\bar{Y}_{i.} = \frac{\sum_{j=1}^n Y_{ij}}{n}, \quad \bar{Y}_{..} = \frac{\sum_{i=1}^k \sum_{j=1}^n Y_{ij}}{nk} \quad \text{حيث ان:}$$

ويمكن التعبير عن مجموعات المربعات في النظرية باستعمال الرموز الآتية:

$$1) SST = \sum_{i=1}^k \sum_{j=1}^n (Y_{ij} - \bar{Y}_{..})^2$$

$$2) SSR = n \sum_{i=1}^k (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$$3) SSE = \sum_{i=1}^k \sum_{j=1}^n (Y_{ij} - \bar{Y}_i.)^2$$

حيث أن SST يمثل مجموع المربعات الكلي Total Sum of Squares

SSR يمثل مجموع مربعات العينات او الصفوف او التأثيرات

Treatment Sums of Squares or between groups

وأن SSE يمثل مجموع مربعات الخطأ

Error Sums of Squares or within groups

ويمكن ملاحظة أن SSE يمكن إيجاده من خلال المعادلة التالية:

$$SSE = SST - SSR$$

ولهذا فجدول تحليل التباين يمكن اختصاره كما يلي:

(ANOVA) Table جدول تحليل التباين الأحادي

مصدر التغير S.V. source of variation	مجموع المربعات ss sum of squares	درجات الحرية df Degrees of freedom	متوسط المربعات mss mean squares	قيمة F المحسوبة Computed F
Between groups بين المجموع (بين العينات)	SSR	K - 1	$MSR = \frac{SSR}{k-1}$	$= \frac{MSR}{MSE}$
Within groups (errors) داخل المجموع (داخل العينات) أو الخطأ	SSE	K(n - 2)	$MSE = \frac{SSE}{K(n-2)}$	
مجموع المربعات الكلي	SST	(Kn - 1)		

أما حساب مجموع المربعات بمساعدة الحاسبة اليدوية فيمكن إيجاده كما يلي:

$$SST = \sum_{i=1}^k \sum_{j=1}^n Y_{ij}^2 - nkY^2..$$

$$SS = n \sum_{i=1}^k Y_i^2 - nkY^2..$$

$$SSE = SST - SSR$$

مثال (2)

استخدمت اربع طرق مختلفة في اربعة شعب من الصف الثاني الابتدائي لتعليم الطلبة جداول الضرب فكانت النتائج كما في الجدول التالي. اختبر فيما لو كان هناك فروق في الطرق المختلفة.

جدول المشاهدات

طرق التدريس	المشاهدات
I	7 6 8 5 9 7
II	8 9 10 7 8 6
III	7 8 10 5 6 3
IV	8 6 5 4 9 4

الحل

اذا فرضنا ان المشاهدات تمثل اربع عينات عشوائية مستقلة عن بعضها البعض وانها اخذت من اربع مجتمعات مستقلة عن بعضها البعض وهي تتوزع توزيع طبيعي ذي المعدلات او المتوسطات التالية $\mu_1, \mu_2, \mu_3, \mu_4$ على التوالي والتباين للعينات متساوي ويساوي σ^2 المطلوب: اختبار فرضية العدم التالية:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

ضد الفرضية البديلة:

$$H_1 : \text{على الاقل احد المتوسطات مختلف}$$

باستخدام مستوى المعنوية $\alpha = 0.05$.

$$k = 4, n = 6$$

فلدينا هنا

$$SST = (7^2 + 6^2 + \dots + 4^2) - (6)(4)(6.875)^2$$

$$= 1232 - 1134.375 = 97.625$$

$$SSR = 6(7^2 + 8^2 + \dots + 6.5^2 + 6^2) - (6)(4)(6.875)^2$$

$$= 1147.5 - 1134.375 = 13.125$$

$$SSE = 97.625 - 13.125 = 84.5$$

ولهذا فجدول تحليل التباين (ANOVA) كما يلي:

قيمة F المحسوبة	متوسط المربعات	درجات الحرية	مجموع المربعات	مصدر التغير
	mss	df	ss	S.V.
1.04	4.375	3	13.125	بين المجموع أو العينات
	4.225	20	84.5	داخل المجموع (داخل العينات) أو الخطأ
F0.05 (3,20)=3.10 الجدولية		23	97.625	الكلية

وبما ان قيمة F المحسوبة* (1.04) اصغر من قيمة F الجدولية (3.10) فإننا لا نرفض فرضية العدم H_0 والتي تقول ان المتوسطات جميعها متساوية او التأثيرات متساوية وتساوي صفر. وهذا يعني لا توجد هناك فروق معنوية بين متوسطات طرق التدريس المختلفة.

مثال (3)

أوجد جدول تحليل التباين الأحادي واختبر الفرضية:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1: \text{على الأقل أحد المتوسطات مختلف}$$

للملاحظات في الجدول التالي:

(طرق تدريس مختلفة) المجاميع أو العينات	المشاهدات Y_{ij} (درجات الطلبة)					
A	7	10	9	8	10	10
B	5	6	4	8	7	6
C	8	9	10	6	3	6
D	5	6	4	3	7	5

الحل

نحسب جدول تحليل التباين الأحادي كما يلي:

$$K = 4$$

$$\bar{Y}_1 = \frac{7+10+9+8+10+10}{6} = 9$$

$$\bar{Y}_2 = \frac{5+6+4+8+7+6}{6} = 6$$

$$\bar{Y}_3 = \frac{8+9+10+6+3+6}{6} = 7$$

$$\bar{Y}_4 = \frac{5+6+4+3+7+5}{6} = 5$$

$$\bar{Y}_{..} = \frac{7+10+\dots+5}{24} = \frac{162}{24} = 6.75$$

$$SST = \sum \sum Y_{ij}^2 - nk\bar{Y}_{..}^2$$

$$= (7+10+\dots+7+5) - (6)(4)(6.75)^2$$

$$= 1206 - 1093.5 = 112.5$$

$$SSR = n \sum \bar{Y}_i^2 - nk\bar{Y}_{..}^2$$

$$SSR = 1146 - 1093.5 = 52.5$$

$$SSE = SST - SSR$$

$$= 112.5 - 52.5 = 60$$

وأخيرا فإن جدول تحليل التباين الأحادي ANONA Table

قيمة F المحسوبة	متوسط المربعات	درجات الحرية	مجموع المربعات	مصدر التغير
	mss	df	ss	S.V.
5.833	17.5	3	52.5	بين المجموع أو العينات
	3	20	60	داخل المجموع (داخل العينات) أو الخطأ
قيمة F الجدولية $F_{0.05}(3,20)=3.10$		23	112.5	الكلي

اختبار تساوي المتوسطات هو اختبار الفرضية التالية:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

H_1 : (على الأقل احد المتوسطات مختلف عن الباقي)

وعند تحديد مستوى المعنوية ولتكن $\alpha = 0.05$ تتم مقارنة F المحسوبة ($F=5.833$) مع F الجدولية والتي تساوي $F_{0.05}(3,20) = 3.10$. وحيث ان F المحسوبة اكبر من F الجدولية لهذا نرفض الفرضية H_0 والتي تنص على تساوي المتوسطات، أي انه على الأقل احد المتوسطات مختلف عن الباقي.

7-4 المقارنات المتعددة Multiple Comparisons:

يعتبر تحليل التباين Analysis of Variance من اهم واقوى الطرق الإحصائية لاختبار تساوي عدة متوسطات للمجتمعات او المجموع، وفي حالة رفضنا للفرضية التي تنص على تساوي المتوسطات، فعند ذلك نقول هناك على الأقل احد المتوسطات مختلف او على الأقل متوسطين غير متساويين وهنا لا نعرف ايا من هذه المتوسطات غير متساوية. وبذلك فإن المقارنات المتعددة ستساعدنا للإجابة على هذا التساؤل.

وهناك حالات عديدة من الضروري جدا معرفة ايا من المتوسطات حقق اعلى فروق معنوية مع باقي المتوسطات في حالة كونه اقل متوسط او اكبر متوسط حسب التجربة او حسب الدراسة، فمثلا اختيار عدة انواع من الاسمدة وما هو افضل سماد

حقوق اعلى كمية انتاج فهنا يفضل السماد الذي يعطى اعلى كمية انتاج او يتم اختيار دواء من مجموعة من الادوية تم إعطاء كلا منها الى مجموعة من المرضى وتحديد عدد ايام الشفاء من مرض معين فهنا يتم اختيار الدواء الذي متوسطه اقل متوسط من الايام اللازمة للشفاء وما الى ذلك من الامثلة على التجارب او الدراسات التي تفضل نوع معين من بين عدة انواع. وهناك عدة اختبارات تستخدم للاجابة على مثل هذه الاسئلة لاختيار احد الانواع او المجاميع وسوف نختار احد اهم هذه الاختبارات وهو اختبار يسمى اختبار شفي Scheffe's Test.

والصيغة العامة لهذا الاختبار هي:

$$CR_s = \sqrt{(a-1)F(df_A, df_{error})} \cdot \sqrt{\frac{2MSE}{N}}$$

حيث ان:

- (a - 1) تمثل عدد مستويات المتغير A مطروح منها واحد.
 N تمثل عدد المشاهدات في كل مستوى من مستويات المتغير.
 MSE هو متوسط مربعات الخطأ والذي تم إيجاده باستخدام جدول تحليل التباين السابق ذكره.
 أما $F(df_A, df_{error})$ فتمثل F الجدولية على درجات الحري.

مثال (4)

تطبيق هذا الاختبار على بيانات المثال (3) السابق لتحديد افضل طريقة تدريس.

الحل

من خلال ايجاد المتوسط الذي يحقق اعلى فروق معنوية مع غيره من المتوسطات نحتاج لحساب قيمة الاختبار CR_s بمستوى معنوية (0.05) أي ان $(\alpha = 0.05)$.

$$CR_{s,0.05} = \sqrt{(a-1)F(df_A, df_{error})} \cdot \sqrt{\frac{2MSE}{N}}$$

حيث ان:

$$CRS_{0.05} = \sqrt{(4-1)F(4-1,20)} \cdot \sqrt{\frac{2(3)}{6}}$$

$$= \sqrt{3(3.10)} \cdot \sqrt{\frac{6}{6}}$$

$$CRS_{0.05} = \sqrt{9.3} = 3.0495 = 3.05$$

ومن ثم يتم ايجاد الفروق بين المتوسطات بعد ترتيب هذه المتوسطات تصاعديا وكما يلي، وبعدها نقارن الفرق بين المتوسطات وبين قيمة اختبار شفي بمستوى $\alpha = 0.05$ والتي تساوي $(CRS_{0.05}=3.05)$ فلكل فرق معنوي نضع (*) والغير معنوي نضع (-) وكما يلي:

	A(9)	C(7)	B(6)	D(5)
A(9)	-	-	-	*
C(7)	-	-	-	-
B(6)	-	-	-	-
D(5)	-	-	-	-

ولهذا فإن افضل طريقة هي الطريقة (A) وهي المسؤولة عن الفرق المعنوي مع الطريقة (D) ولهذا نختار الطريقة A كأفضل طريقة.

7-5 اختبار تساوي عدة تباينات The Test of the equality of variances

من الأمور المهمة والأساسية في تحليل التباين (Analysis of Variance) هو اختبار تساوي التباينات للمجاميع المختلفة التي هي تحت الدراسة وهناك عدة اختبارات واحد اهم هذه الاختبارات هو اختبار بارتلت Bartlett's Test. والفرضيات الخاصة بهذا الاختبار هي:

الفرضية المراد اختبارها H_0 هي ان التباينات متساوية أي ان:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots \sigma_K^2$$

اما الفرضية البديلة H_1 فهي:

على الاقل عدم تساوي اثنان من التباينات او على الاقل احد التباينات مختلف:

H_1

ولحساب هذا الاختبار نحتاج الى ما يلي:

$$B = 2.3026 \frac{M}{C}$$

الصيغة الخاصة بالاختبار

حيث ان M و C هما كما يلي على التوالي:

$$M = (N - K) \log s_p^2 - \sum (n_i - 1) \log s_i^2$$

$$C = 1 + \frac{1}{3(k-1)} \left[\sum \frac{1}{n_i - 1} - \frac{1}{N - K} \right]$$

وأن:

n_1, n_2, \dots, n_k : تمثل حجوم العينات بحيث ان:

$$\sum n_i = N$$

$$s_p^2 = \frac{\sum (n_i - 1) s_i^2}{N - K}$$

وكذلك فإن:

بحيث ان s_i^2 هو تباينات العينات وأن K هو عدد العينات.

قيمة B تقترب من توزيع " مربع كاي " أو χ^2 بدرجة حرية $v = k - 1$ وبمستوى المعنوية α . وفي حالة كون B المحسوبة اكبر من قيمة B الجدولية فهذا يعني أن التباينات غير متساوية والعكس صحيح.

مثال
(5)

استخدم بيانات المثال الثاني لتطبيق اختبار تساوي التباينات.

الحل

باستخدام الصيغة الخاصة باختبار بارتل $Bartlett's Test$

$$S_1^2 = 2 \quad S_2^2 = 2 \quad S_3^2 = 5.9 \quad S_4^2 = 4.4 \text{ لدينا}$$

$$B = 2.3026 \text{ (M/C)} \quad n_1=6, n_2=6, n_3=6, n_4=6 \quad N=24 \quad K=4$$

$$M = (N - K) \log S_p^2 - \sum (n_i - 1) \log S_i^2$$

$$S_p^2 = \frac{\sum (n_i - 1) S_i^2}{N - K} = \frac{5 \times 2 + 5 \times 2 + 5 \times 5.9 + 5 \times 4.4}{24 - 4} = \frac{71.5}{20} = 3.575$$

$$\begin{aligned} \sum (n_i - 1) \log S_i^2 &= 5 \times \log 2 + 5 \times \log 2 + 5 \times \log 5.9 + 5 \times \log 4.4 \\ &= 1.51 + 1.51 + 3.854 + 3.217 = 10.1 \end{aligned}$$

$$M = 20 \log 3.575 - 10.091$$

$$= 11.1 - 10.1 = 1$$

$$\begin{aligned} C &= 1 + \frac{1}{3(k-1)} \left[-\frac{1}{n_i - 1} - \frac{1}{N - K} \right] \\ &= 1 + \frac{1}{9} \left[\frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5} - \frac{1}{20} \right] \\ &= 1 + \frac{1}{9} \left[\frac{4}{5} - \frac{1}{20} \right] = 1 + \frac{1}{9} [0.8 - 0.05] \end{aligned}$$

$$C = 1 + 0.0833 = 1.083$$

$$B = 2.3026 \times \frac{1}{1.083}$$

$$= 2.3026 \times 0.923$$

$$B = 2.126$$

وبما ان قيمة B المحسوبة (2.126) اقل من قيمة B الجدولية ($\chi_{0.05,3}^2 = 7.8$) لذا لا نرفض الفرضية H_0 التي تقول ان التباينات متساوية.

**مثال
(6)**

تطبيق اختبار بار تلت على تباينات المثال (3) لتحديد فيما اذا كانت التباينات متساوية أم لا.

الجدول الصيغة الخاصة باختبار بارلتت Bartlett's Test هو:

$$B = 2.3026 (M/C)$$

ومن معلومات المثال لدينا: $n_1=n_2=n_3=n_4=6$

$$N = \sum ni = N = 24 \quad K = 4$$

$$S_1^2 = 1.6 \quad S_2^2 = 2 \quad S_3^2 = 6.4 \quad S_4^2 = 2$$

$$M = (N - K) \log S_p^2 - \sum (ni - 1) \log S_i^2$$

$$S_p^2 = \frac{\sum (ni - 1) S_i^2}{N - K} = \frac{5 \times 1.6 + 5 \times 2 + 5 \times 6.4 + 5 \times 2}{24 - 4} = \frac{60}{20} = 3$$

$$\sum (ni - 1) \log S_i^2 = 5 \times 0.204 + 5 \times 0.3 + 5 \times 0.806 + 5 \times 0.3 = 8.05$$

$$M = 20 \log 3 - 8.05 = 20 \times 0.477 - 8.05$$

$$= 9.542 - 8.05$$

$$M = 1.49$$

$$C = 1 + \frac{1}{3 \times 3} \left[- \frac{1}{ni} - \frac{1}{N - K} \right]$$

$$= 1 + \frac{1}{9} \left[\frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{1}{5} - \frac{1}{20} \right] = 1.083$$

$$B = 2.3026 \times \frac{1.49}{1.083}$$

$$B = 2.3026 \times 1.376$$

$$B = 3.168$$

وبما ان قيمة B المحسوبة ($B=3.168$) اقل من قيمة B الجدولية ($\chi_{0.05,3}^2 = 7.8$) لذا لا نرفض الفرضية H_0 التي تنص على ان التباينات متساوية.

تطبيقات SPSS:

ويمكن إيجاد تحليل التباين لمتغير واحد one - way كما يلي: نختار من القائمة الرئيسية Analyze ومن هذه القائمة نختار الخيار Means Compare ومن هذا الخيار نحدد الخيار - One Way ANOVA وبعد تحديد المتغيرات المعتمد والذي يمثل القياسات والمستقل والذي يمثل المجاميع. ومن هذه الشاشة يمكن تحديد خيار المقارنات المتعددة من خيار Post Hoc. واختبار التباينات من خلال الخيار options. وبعد تحديد هذه الخيارات نقر على المفتاح ok للتنفيذ.

أسئلة

الفصل السابع

1- الجدول التالي يمثل نتائج ثلاث مجاميع مختلفة من طلبة جامعة عمان الأهلية تم تدريسهم مادة الرياضيات من قبل ثلاثة أساتذة.

الأساتذة

A	B	C
50	60	70
60	65	80
55	75	85
62	80	82
64	82	84
70	74	86
40	70	90
58	78	93

المطلوب:

- هل هناك فروق معنوية بين متوسطات المجاميع الثلاث؟ أوجد جدول تحليل التباين (ANOVA) واختبر بمستوى معنوية 1%.
- اختبر تجانس التباينات باستخدام اختبار بارتلت.
- هل يوجد هناك أستاذ متميز أم لا، استخدم المقارنات المتعددة وذلك من خلال اختبار شفي بمستوى 1%.

2- الجدول التالي يمثل نتائج تجربة زراعية على أربعة أنواع من السماد تم استخدامها لمحصول الذرة الصفراء ولمجموعة من المساحات المتساوية بالمساحة والخصوبة.

أنواع السماد

A	B	C	D
10	4	10	15
8	5	11	16
6	6	12	17
7	3	14	18
9	4	13	19
5	5	9	20
7	3	13	18
9	5	15	19

المطلوب:

- أوجد هل هناك فروق معنوية بين متوسطات الأنواع الأربعة من الأسمدة؟
أوجد جدول تحليل التباين (ANOVA) واختبر بمستوى معنوية 5%.
- اختبر تجانس التباينات باستخدام اختبار بار تلت بمستوى معنوية 5%.
- هل هناك نوع يفضل على غيره يصلح للتصنيع يختلف معنوياً بالإنتاج عن الباقي، استخدم اختبار شفي (Scheff's Test) بمستوى 5%.
- 3- قارن بين التأثيرات A وB وC مستخدماً جدول تحليل التباين:

A : 5 6 7

التأثيرات B : 3 4 6 7

C : 4 5 6 6 7 8

- 4- البيانات التالية تمثل إنفاق الطاقة للسنة الماضية لعدد من الأسر في أربعة مناطق مختلفة:

المنطقة الأولى : 7 9 11 13

المنطقة الثانية : 7 8 9 10 12

المنطقة الثالثة : 10 13 13 14 15

المنطقة الرابعة : 12 12 13 15 17 18

- قارن بين تأثير المناطق المختلفة.

- حدد أي من التأثيرات هو المسبب للاختلاف.

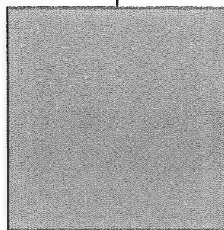
5- أكمل جدول تحليل التباين التالي:

S.V.	S.S.	d.f.	M.S.	F	P
بين التأثيرات		5	2503.02		
داخل التأثيرات		54			
الكلي	13484.583				

6- أكمل جدول تحليل التباين التالي الخاص بمقارنة ثلاث تأثيرات وأن عدد المفردات الكلية 60:

S.V.	S.S.	d.f.	M.S.	F	P
بين التأثيرات				1.28	
داخل التأثيرات			43.793		
الكلي	13484.583				

الرجاء



8

الفصل الثامن

الطرق

اللامعلمية

**Non-Parametric
Methods**

8-1 مقدمة

8-2 اختبار "ولكوكسن"

8-3 اختبار "مان - ويتني"

8-4 اختبار "ولكوكسن" للفرق المزدوج

8-5 اختبار "كروسكل - والس"

8-6 اختبار معامل الارتباط

8-7 اختبار حسن المطابقة

8-8 اختبار الاستقلالية

الفصل الثامن

8

الفصل الثامن الطرق الالامعلمية

Non-Parametric Methods

8-1 مقدمة Introduction:

كما لاحظنا أن جميع الأساليب الإحصائية التي استخدمت في هذا الكتاب تعتمد على حقيقة مهمة وهي أن التوزيع المستخدم هو التوزيع الطبيعي، سواء كان توزيع المجتمع طبيعيا أو إذا كان حجم العينة كبيرا فإن التوزيع التقريبي هو التوزيع الطبيعي.

في هذا الفصل يتم التعرف على بعض الأساليب الإحصائية والتي لا تعتمد على مبدأ التوزيع الطبيعي، بل إن التوزيعات المستخدمة قد تكون أي من التوزيعات المتماثلة أو غير ذلك أو أن يكون التوزيع غير معلوما لبعض من الحالات الأخرى، وبذلك فإن الطرق الإحصائية التي تعرض في هذا الفصل تدخل تحت اسم الطرق الالامعلمية Non-Parametric Methods على نظير أن الطرق السابقة يطلق عليها اسم الطرق المعلمية Parametric Methods حيث إنها تعتمد على معالم أو مؤشرات التوزيع المستخدم.

سنعرض في هذا الفصل بعض من هذه الطرق الالامعلمية والتي لها أهمية كبيرة في الجانب التطبيقي لطلبة الإدارة والاقتصاد ولن يتم التطرق للطرق الأخرى والتي يكون تطبيقها خارج عن نطاق هذا الكتاب، الطرق المعروضة في هذا الفصل لها ترتيب مشابه للطرق التي تم عرضها في الفصول السابقة من هذا الكتاب تحت اسم طرق الاختبار، وبالشكل التالي:

8-2 اختبار "واكسون" The Wilcoxon Signed-Rank Test:

تستخدم هذه الطريقة للتطبيقات التي يكون فيها التوزيع متماثل بحيث يمكن

قسمته إلى نصفين متساويين في نقطة المنتصف المتوسط أو الوسيط من دون الاعتماد على شرط أن يكون التوزيع طبيعياً، وأن خطوات هذه الطريقة هي:

الخطوة الأولى: تحديد الفرضية المراد اختبارها H_0 والفرضية البديلة H_1 في أن الوسط أو الوسيط يساوي قيمة معينة ولتكن μ_0 أو Med_0 بأحد الأشكال الثلاثة السابق ذكرها، أي أن $H_0 : \mu = \mu_0$ وأن الفرضية البديلة قد تكون أي من الأشكال التالية:

$$H_1 : \mu < \mu_0 \quad \text{أو} \quad H_1 : \mu > \mu_0 \quad \text{أو} \quad H_1 : \mu \neq \mu_0$$

الخطوة الثانية: تحديد مستوى المعنوية α .

الخطوة الثالثة: حساب إحصاء الاختبار والتي يرمز لها بالرمز (+) ويتم حسابها كما يلي:

أولاً إيجاد قيم الجدول التالي:

الترتيب بالإشارة R_i	رتب الفرق المطلق	الفريق المطلق $ D_i $	الفرق $D_i = X_i - \mu_0$	القيم X_i
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮

وبالتالي علينا إيجاد الفرق بين القيم X_i والقيمة المحددة μ_0 وهذا الفرق هو D_i ثم إيجاد القيم المطلقة $|D_i|$ وبعد ذلك يتم إعطاء رتب للقيم المطلقة بحيث تعطى الرتبة 1 للقيمة الأصغر والرتبة 2 للقيمة التي بعدها وهكذا، وأخيراً يتم إعطاء الإشارة الأصلية لرتب الفرق لنحصل بذلك على R_i .

وإن إحصاء الاختبار ω^+ تساوي مجموع الرتب الموجبة.

الخطوة الرابعة: تحديد منطقة الرفض ويتم ذلك بملاحظة مستوى المعنوية α واستخدام الجداول الخاصة بالقيم الحرجة حيث أن الشكل التالي يوضح الحالة.

اختيار من جهة اليسار		اختيار من جهة اليمين	
ω_u^+	منطقة عدم	ω_u^+	منطقة عدم
منطقة رفض H_0	الرفض H_0	الرفض H_0	منطقة رفض H_0

اختيار من جهة اليسار		اختيار من جهة اليمين	
ω_L^+	منطقة عدم	ω_u^+	منطقة عدم
منطقة رفض H_0	رفض H_0	منطقة رفض H_0	منطقة رفض H_0

الخطوة الخامسة: مقارنة القيمة المستخرجة مع القيم الجدولة ومن ثم اتخاذ القرار برفض او عدم رفض H_0 .

ويلاحظ من اتباع الخطوات أعلاه ومقارنتها بالخطوات التي تم اتباعها سابقا أن هذا الاختبار مشابه لاختبار متوسط واحد، غير هذا الاختبار اكثر قوة في حالة أن التوزيع المستخدم غير التوزيع الطبيعي.

إذا كان حجم العينة كبيرا يمكن استخدام التوزيع الطبيعي كتوزيع تقريبي لإجراء عملية الاختبار السابقة كما يلي:

يمكن القول أن ω^+ لها توزيع طبيعي تقريبي بالوسط والتباين التاليين:

$$\mu_{\omega^+} = \frac{n(n+1)}{4}$$

$$\sigma_{\omega^+}^2 = \frac{n(n+1)(2n+1)}{24}$$

وبذلك فإن اختبار الفرضية $H_0: \mu = \mu_0$ نستند على استخدام إحصاءة الاختبار Z_0 والتي لها توزيع طبيعي معياري حيث أن:

$$Z_0 = \frac{\omega^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

ويتم الاختبار كما سبق ذكره باستخدام التوزيع الطبيعي.

مثال (1)

افترض البيانات التالية واختبر الفرضية التي تنص على أن الوسط هو 01 باستخدام $\alpha = 0.05$

6.2 ، 12.1 ، 8.5 ، 15.8 ، 17.7 ، 9.6 ، 4.0 ، 25.5 ، 10.3 ، 13.9

الحل

بما أن هذا الاختبار يخص عينة واحدة وباتباع خطوات العمل السابق ذكرها نحدد أولاً الفرضية H_0 والفرضية البديلة H_1 بالشكل:

$$H_0 : \mu = 10$$

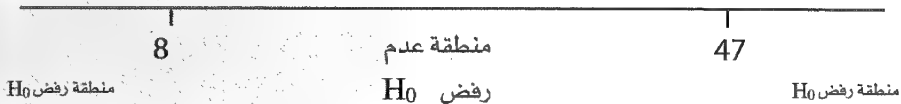
$$H_1 : \mu \neq 10$$

وإن إحصاء الاختبار w^+ هي مجموع الرتب الموجبة بعد عمل الجدول التالي:

X_i العنصر	$X_i - \mu_0$	$ D_i $	رتب $ D_i $	R_i
6.2	-3.8	3.8	5	-5
12.1	2.1	2.1	4	4
8.5	-1.5	1.5	3	-3
15.8	5.8	5.8	7	7
17.7	7.7	7.7	9	9
9.6	-0.4	0.4	2	-2
4.0	-6.0	6.0	8	-8
25.2	15.2	15.2	10	10
10.3	0.3	0.3	1	1
13.9	3.9	3.9	6	6

وبذلك فإن $w^+ = 37$.

وباستخدام $\alpha = 0.05$ ومن جدول في الملحق نجد أن منطقة الرفض تتحدد بالشكل التالي حيث أن $w^+_{L} = 8$ و $w^+_{U} = 47$.



وبما أن w^+ تقع ضمن منطقة عدم رفض H_0 لذلك لا نرفض H_0 ، أي أن الوسط يساوي 10.

أما باستخدام التوزيع الطبيعي كتوزيع طبيعي لعمل الاختبار فلدينا:

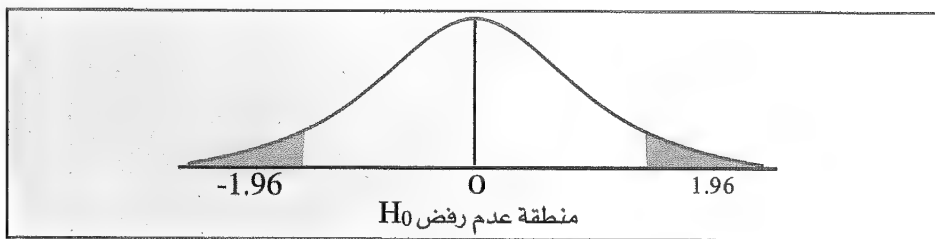
$$\mu_{w^+} = \frac{n(n+1)}{4} = \frac{(10)(11)}{4} = 27.5$$

$$\sigma_{w^+}^2 = \frac{n(n+1)(2n+1)}{24} = \frac{(10)(11)(21)}{24} = 96.25$$

أي أن إحصاءة الاختبار Z_0 هي:

$$Z_0 = \frac{w^+ - \mu_{w^+}}{\sigma_{w^+}} = \frac{37 - 27.5}{\sqrt{96.25}} = 0.968$$

وباستخدام التوزيع الطبيعي المعياري وان $\alpha = 0.05$ نجد أن منطقة الرفض تتحدد بالشكل:



وبما أن إحصاءة الاختبار تقع ضمن منطقة عدم الرفض، لذلك لا نرفض H_0 أي أن الوسط يساوي 10.

تطبيقات SPSS:

ويمكن إيجاد قيمة هذا الاختبار كما يلي:

نختار من القائمة الرئيسية إلى SPSS الخيار Analyze ومن هذه القائمة نختار Nonparametric tests ومن هذه القائمة نختار الخيار 2 Related Samples في حالة الاختبار بين عيتين وهنا نحدد كل عينة ضمن متغير ونحدد الخيار Wilcoxon ونقر على الزر ok للتنفيذ.

8-3 اختبار " مان- ويتني " The Mann-Whitney Test:

هذه الطريقة تستخدم وبشكل مشابه لما تم عمله في اختبار متوسطين حيث أننا نقوم هنا باختبار الفرق بين متوسطين من مجتمعين باستخدام عيتين عشوائيتين مستقلتين. ويستخدم هذا الاختبار عندما يكون التوزيعان لهما نفس الشكل من دون التطرق لكونهما طبيعيا أو بأي من المسميات الأخرى. وهذه الطريقة تُلخص بالخطوات التالية:

الخطوة الأولى: تحديد الفرضية المراد اختبارها H_0 وهي أن المتوسطان متساويان، أي أن $H_0 : \mu_1 = \mu_2$ ، وأيضا يتم تحديد الفرضية البديلة وقد تكون أي من الأشكال الثلاثة

$$H_1 : \mu_1 < \mu_2 \text{ أو } H_1 : \mu_1 > \mu_2 \text{ أو } H_1 : \mu_1 \neq \mu_2$$

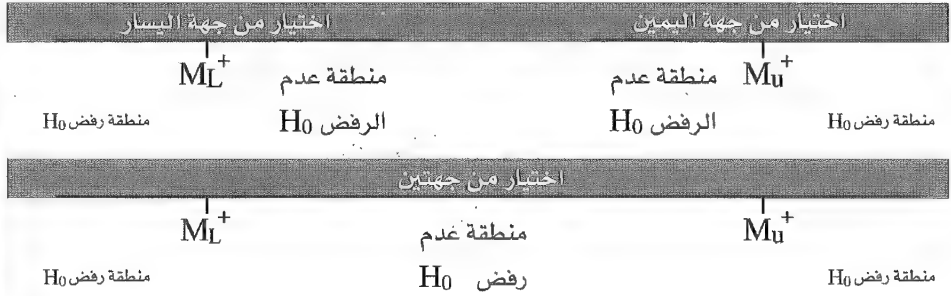
الخطوة الثانية: تحديد مستوى المعنوية α .

الخطوة الثالثة: إيجاد قيمة إحصاء الاختبار والتي يرمز لها بالرمز M وإيجادها يتم باستخدام الجدول التالي:

الترتيب R	القيم للعينة من التوزيع (II)	الترتيب R	القيم للعينة من التوزيع (I)
.	.	.	.
.	.	.	.
.	.	.	.

وذلك يعني إعطاء رتب لقيم العيتين مع بعضهما بحيث تعطى الرتبة 1 للقيمة الأصغر ثم الرتبة 2 للقيمة التي بعدها وهكذا.

أما إحصاء الاختبار M فتمثل مجموع الرتب الخاصة بقيم العينة من التوزيع (I) الخطوة الرابعة: باستخدام مستوى المعنوية α يتم تحديد منطقة الرفض كما يظهر بالشكل التالي:



الخطوة الخامسة: يتم اتخاذ القرار بشأن الفرضية.

أما عندما يكون حجم العينة كبيرا فنستخدم التوزيع الطبيعي حيث أن إحصاء الاختبار M لها توزيع طبيعي تقريبي بالوسط والتباين التاليين:

$$\mu_M = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$\sigma_M^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

ومن ثم يتم اختيار الفرضية كما تم ذكره سابقا في اختبار الفرضيات باستخدام التوزيع الطبيعي.

مثال (2)

في دراسة مقارنة لنوعين من الطباعة، استخدمت عينة عشوائية من 10 من الأوراق المطبوعة بحيث طبعت الخمسة الأوراق الأولى بالنوع A من الطباعة، أما الخمسة الأخيرة فطبعت بالنوع B من الطباعة. البيانات التالية تمثل الوقت اللازم لقراءة الورقة المطبوعة كآلاتي:

108، 99، 101، 122، 95 : النوع A.

120، 112، 115، 102، 110 : النوع B.

اختبر فيما إذا كانت البيانات تعطي الأدلة الكافية على أن الطباعة بالنوع A أسهل في القراءة من الطباعة بالنوع B. استخدم $\alpha = 0.05$.

الحل

لإجراء الاختبار للفرضية $H_0: \mu_1 = \mu_2$ مع الفرضية البديلة $H_1: \mu_1 < \mu_2$ علينا تحويل القيم إلى الرتب كآلاتي :

A. النوع : 1، 10، 3، 2، 5

B. النوع : 6، 4، 8، 7، 9

وبالتالي فإن إحصاء الاختبار $M = 21$ وباستخدام $\alpha = 0.05$ والرجوع إلى الجدول في الملحق نجد أن مناطق الرفض كما في الشكل التالي :

19

منطقة عدم

رفض H_0

وبما أن إحصاء الاختبار تقع ضمن منطقة عدم الرفض، لذلك لا نرفض H_0 ونقول بأن القراءة للطباعة بالنوع A ليست أسهل من القراءة للطباعة بالنوع B.

تطبيقات SPSS:

ويمكن إيجاد هذا الاختبار كما يلي :

نختار Analyze من القائمة الرئيسية إلى SPSS ومن هذه القائمة نختار Nonparametric tests ومن هذه القائمة نختار الخيار 2 Related Samples ومن هذه الشاشة نضع المتغيرات لكل عينة في حقل Test Variable ونضع متغير المجاميع في حقل Grouping ونختار الخيار Mann - Whitney ونقر على (ok).

مثال (3)

هل أن توزيع أعداد الملتحقين من الطلبة العرب حسب الجنسية للجامعات الأردنية متساوي للعام الدراسي الجامعي 1992/1993. باستخدام $\alpha = 0.05$.

الجنسية	الإناث	الذكور
السعودية	18	28
عمان	71	115
العراق	87	162
فلسطين	861	2062
قطر	8	9
الإمارات	36	5
سوريا	70	95
المغرب	11	4
مصر	14	5
لبنان	25	31
السودان	34	36
اليمن	47	554
البحرين	10	3
تونس	3	1
الجزائر	3	1
الكويت	4	4
ليبيا	3	6

الحل

الفرضية المراد اختبارها والبديلة هما

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

ولأجل عمل الاختبار علينا أولاً تحويل القيم إلى رتب كما تم وصفه وبالتالي فإنها تظهر بالشكل التالي:

رتب الإناث: 4.5, 8, 4.5, 4.5, 15, 25, 22, 19, 17, 16, 26, 23.5, 13, 33, 28, 27, 18.

رتب الذكور: 12, 8, 1.5, 1.5, 4.5, 32, 23.5, 21, 10.5, 8, 29, 10.5, 14, 34, 31, 30, 20.

أما إحصاء الاختبار M فهي مجموع الرتب لقيم العينة من التوزيع (1) ولنعتبره

الإناث، لذلك فإن: $M=304$

وبما أن حجم العيتين كبيراً، لا يمكن استخدام الجداول الخاصة بهذا الاختبار كما تم عمله في مثال (2) السابق. بل علينا استخدام الصيغة التقريبية للتوزيع الطبيعي، حيث أن:

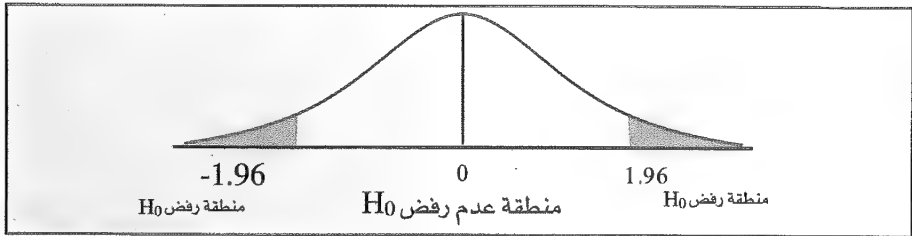
$$\mu_M = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{(17) \times (17 + 17 + 1)}{2} = 297.5$$

$$\sigma_M^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12} = \frac{(17)(17)(17 + 17 + 1)}{12} = 842.9167$$

وبذلك فإن إحصاء الاختبار Z_0 هي:

$$Z_0 = \frac{M - \mu_M}{\sigma_M} = \frac{304 - 297.5}{\sqrt{842.9167}} = \frac{6.5}{29.033} = 0.2239$$

وباستخدام جدول التوزيع الطبيعي المعياري وان $\alpha = 0.05$ نجد أن منطقة الرفض هي بالشكل:



وبما أن إحصاء الاختبار تقع ضمن منطقة عدم الرفض، لذلك لا نرفض H_0 ، أي أن المتوسطان متساويان.

8-4 اختبار "ولكوكسن" للفرق المزدوج:

The Paired Wilcoxon Signed-Rank Test

هذه الطريقة مشابهة لطريقة اختبار الفرق بين متوسطين للبيانات المزدوجة، وتعتمد بشكل أساسي على فرضية أن الفروقات متماثلة كمختصر للقول بأن المتغير

الذي يمثل الفروقات المزدوجة له توزيع متماثل. ويلاحظ أنه وبعد احتساب الفروق تصبح المسألة من كونها اختبار الفرق بين متوسطين إلى اختبار حول متوسط واحد وبذلك نعود إلى اختبار ولكوكس الأول والذي تم عمله في المبحث 2-8 السابق، وتعتمد هذه الطريقة على الخطوات التالية:

الخطوة الأولى: تحديد الفرضية المراد اختبارها H_0 والفرضية البديلة والتي تأخذ أحد الأشكال الثلاثة المتعارف عليها لتكون $H_0: \mu_1 = \mu_2$

أما البدائل فهي: $H_1: \mu_1 < \mu_2$ أو $H_1: \mu_1 > \mu_2$ أو $H_1: \mu_1 \neq \mu_2$

الخطوة الثانية: تحديد مستوى المعنوية α

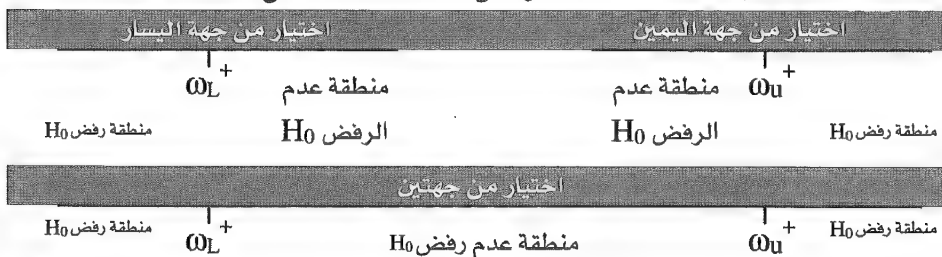
الخطوة الثالثة: إيجاد قيمة إحصاء الاختبار والتي يرمز لها بالرمز ω والتي يتم إيجادها كما يلي وباستخدام الجدول التالي:

X_i	y_i	$d_i = X_i - y_i$	$ d_i $	رتب $ d_i $	الرتب $ d_i $ بالإشارة R_i
.
.
.

وذلك يعني إيجاد الفرق d_i بين القيم X_i والقيم y_i ، أي أن $d_i = x_i - y_i$ ثم إيجاد $|d_i|$ والتي تمثل القيم المطلقة للفروق ومن ثم إعطاء الرتب للقيم $|d_i|$ بحيث تعطى الرتبة (1) للقيمة الأصغر والرتبة (2) للقيمة التي بعدها وهكذا، وأخيراً إعطاء الإشارة الأساسية للفروق لرتب $|d_i|$ لنحصل على القيم R_i .

أما إحصاء الاختبار ω^+ فتمثل مجموع الرتب الموجبة.

الخطوة الرابعة: تحديد منطقة الرفض بالاستعانة بالشكل:



الخطوة الخامسة: اتخاذ القرار بشأن الفرضية.

تطبيقات SPSS:

نختار من القائمة الرئيسية Analyze ومن هذه القائمة نختار 2 Related Samples وبعد تحديد العينات لكل متغير في موقع المتغيرات Test pair (s) list ونختار الاختيار wilcoxon بعدها نقر على الزر ok للتنفيذ.

8-5 اختبار " كروسكل -والس " The Kruskal -Wallis Test:

وهي الطريقة البديلة لجدول تحليل التباين والتي تم ذكرها سابقا حيث يمكن اختبار تساوي عدد من المتوسطات بدون استخدام فرضية التوزيع الطبيعي بل أن هذه الطريقة تعتمد أساسا على كون المتغيرات تحت الدراسة لها نفس الشكل وان العينات المختلفة مستقلة عن بعضها، أما خطوات هذه الطريقة فهي:

الخطوة الأولى: تحديد الفرضية المراد اختبارها H_0 والفرضية البديلة H_1 كآلاتي:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_K$$

على الأقل واحد من هذه المتوسطات مختلف: H_1

الخطوة الثانية: تحديد مستوى المعنوية α

الخطوة الثالثة: إيجاد قيمة إحصاء الاختبار والتي يرمز لها بالرمز H والتي

يمكن إيجادها باستخدام الجدول التالي:

القيم للعينات من المجتمع (I)	الرتب R_1	القيم للعينات من المجتمع (II)	R_2	القيم للعينات من المجتمع (k)	R_k
.
.
.

وذلك يعني إعطاء الرتب R للقيم من المجتمعات المختلفة حيث تعطى الرتبة (1)

للقيمة الأصغر ثم الرتبة (2) للقيمة التي بعدها وهكذا.

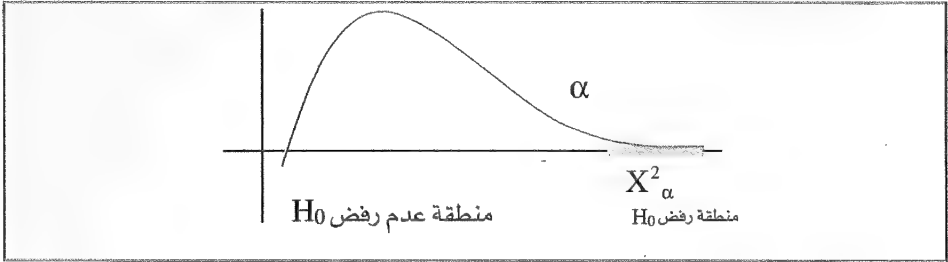
أما إحصاء الاختبار H فهي :

$$H = \frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} - 3(n+1)$$

حيث أن R_1, R_2, \dots, R_K تمثل مجموع الرتب للمفردات في المجتمعات 1, 2, ..., k على التوالي، أما n فتمثل مجموع عدد المفردات.

إحصاء الاختبار H تتبع توزيع "مربع كاي" X^2 بدرجات حرية $(k-1)$ حيث k يمثل عدد المجتمعات.

الخطوة الرابعة: تحديد منطقة الرفض كما في الشكل التالي :



الخطوة الخامسة: اتخاذ القرار بشأن اختبار الفرضية.

مثال (5)

البيانات التالية تمثل توزيع الجهاز الأكاديمي لبعض من الجامعات الأردنية حسب الدرجة العلمية للعام 1993/1992 .

الدرجة العلمية	دكتوراه	ماجستير	دبلوم عالي	بكالوريوس
الجامعة الأردنية	665	146	26	46
جامعة اليرموك	449	110	3	47
جامعة عمان الأهلية	91	14	0	0

هل أن توزيع الجهاز الأكاديمي نفسه للجامعات المختلفة، باستخدام $\alpha = 0.01$.

علينا أولاً تحديد الفرضية المراد اختبارها H_0 والفرضية البديلة لتكون

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

على الأقل أحد هذه الأوساط مختلف H_1 :

حيث أن μ_1, μ_2, μ_3 هي الأوساط للأعداد في الجامعة الأردنية، جامعة اليرموك، جامعة عمان الأهلية على التوالي.

ولحساب إحصاء الاختبار H ، علينا تحويل القيم إلى رتب كما مبين أدناه:

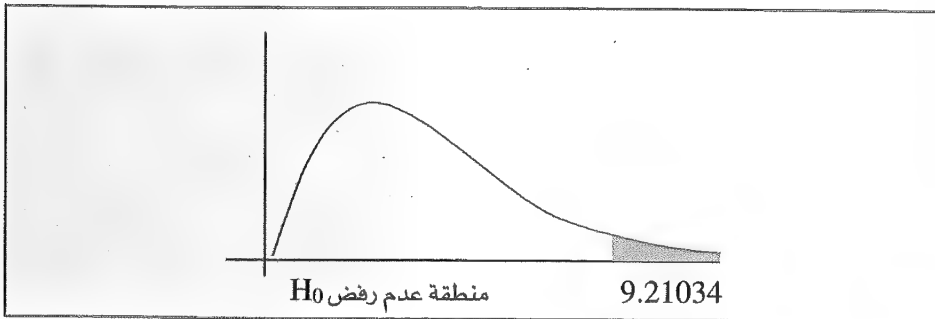
الرتب للجامعة الأردنية	12	10	5	7
الرتب لجامعة اليرموك	11	9	3	6
الرتب لجامعة عمان الأهلية	8	4	1.5	1.5

وأن $\sum R_1 = 34$, $\sum R_2 = 29$, و $\sum R_3 = 15$ وبذلك فإن :

$$H = \frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} - 3(n+1)$$

$$= \frac{12}{(12)(13)} \left[\frac{(34)^2}{4} + \frac{(29)^2}{4} + \frac{(15)^2}{4} \right] - 3(13) = 3.7308$$

وباستخدام $\alpha = 0.01$ وبالرجوع إلى جدول توزيع X^2 بدرجات حرية 2 نجد أن $X^2_{(2, 0.01)} = 9.21034$ وبذلك فإن منطقة الرفض ستظهر بالشكل التالي:



وبما أن إحصاءة الاختبار تقع ضمن منطقة عدم الرفض، لذلك لا نرفض H_0 ، أي أن الأوساط متساوية.

نطبيقات SPSS:

لايجاد هذا الاختبار كروسكل - والس نقوم بالخطوات التالية:

نختار الخيار Analyze من القائمة الرئيسية ومن هذه القائمة نختار Nonparametric Tests ومن هذه القائمة نختار K Independent Samples. بعدها نضع المتغيرات التي تمثل العينات في Test - Variable ونضع متغير المجاميع في حقل Grouping ونعرف المجاميع وبعدها ننقل على الزر ok للتنفيذ.

8-6 اختبار معامل الارتباط:

Spearman's Non-Parametric Test for Rank Correlation

ويخص هذا الاختبار معامل الارتباط، حيث نستطيع من خلاله معرفة فيما إذا كانت المتغيرات أو الظواهر مرتبطة ببعضها أم لا. ويعتمد على مبدأ العشوائية في اختيار المفردات وأن التوزيعات المستخدمة هي توزيعات مستمرة، أما خطوات هذه الطريقة فهي:

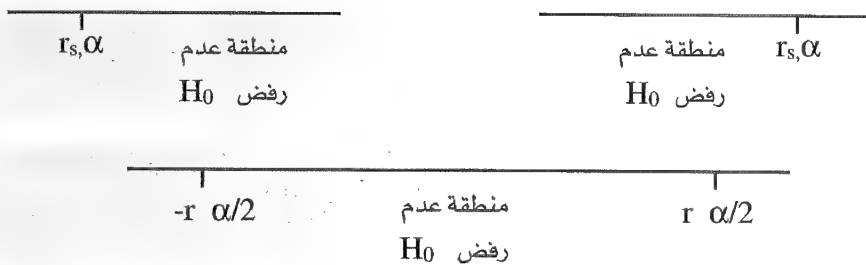
الخطوة الأولى: تحديد الفرضية المراد اختبارها $H_0: \rho = 0$ والفرضية البديلة بأحد الأشكال:

$$H_1: \rho > 0 \quad \text{أو} \quad H_1: \rho < 0 \quad \text{أو} \quad H_1: \rho \neq 0.$$

الخطوة الثانية: تحديد مستوى المعنوية α .

الخطوة الثالثة: حساب قيمة إحصاءة الاختبار وتمثل r وهو معامل ارتباط الرتب السابق ذكره في الفصل الخاص بالانحدار والارتباط.

الخطوة الرابعة: تحديد منطقة الرفض والنقاط الحرجة كما يلي باستخدام الجدول في الملحق:



الخطوة الخامسة: اتخاذ القرار بشأن الفرضية.

مثال (6)

هل هناك علاقة بين الدخل والمبالغ المقرضة من البنك والتي تظهر بياناتها لعدد من العملاء كآلاتي. اختبر باستخدام $\alpha = 0.01$.

الدخل: 16100, 12700, 9300, 41800, 13700, 18200, 22100, 83600, 8900, 14800

المبالغ المقرضة: 1800, 6100, 3200, 0, 3700, 5500, 3300, 500, 4800, 4300

الحل

لاختبار معامل الارتباط، علينا أولاً حساب قيمة معامل ارتباط الرتب لهذه البيانات والتي تتم باتباع الخطوات التالية، حيث علينا تحويل القيم إلى رتب بافتراض أن x_i يمثل الدخل وأن $R(x_i)$ تمثل رتبها وكذلك بافتراض أن y_i يمثل المبالغ المقرضة وأن $R(y_i)$ تمثل رتبها وبالتالي فإن:

$R(x_i)$:	5	1	10	8	7	4	9	2	3	6
$R(y_i)$:	7	8	2	5	9	6	1	4	10	3
$[R(x_i) - R(y_i)]^2$:	4	49	64	9	4	4	64	4	49	9

وبذلك فإن معامل ارتباط الرتب r هو :

$$r = 1 - \frac{6 \sum [R(x_i) - R(y_i)]^2}{n(n^2 - 1)}$$

$$= 1 - \frac{(6)(260)}{(10)(99)} = -0.576$$

والذي يشير إلى أن العلاقة سالبة متوسطة.

أما عن اختبار الارتباط فيتم بافتراض الفرضية H_0 والفرضية البديلة H_1 كآلاتي :

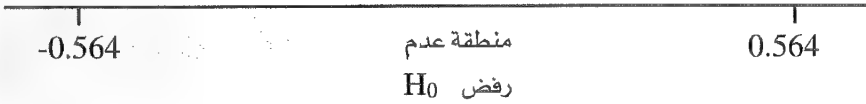
$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

وباستخدام $\alpha = 0.01$ فإن القيمة المجدولة هي :

$$r_{a/2} = 0.564 \text{ و } -r_{a/2} = -0.564$$

وبذلك فإن مناطق الرفض هي بالشكل :



وبما أن إحصاء الاختبار تقع ضمن منطقة الرفض، لذلك نرفض H_0 ، أي أن معامل الارتباط يختلف عن الصفر وذلك يعني وجود ارتباط معنوي بين الدخل والمبالغ المقرضة.

8-7 اختبار حسن المطابقة: Goodness of Fit Test

هذا الاختبار يختلف عن الاختبارات التي سبق ذكرها في أنه لا يبحث في الاختبارات الخاصة بالمؤشرات مثل الوسط أو الوسيط أو التباين، بل أن هذا الاختبار يختص باختبار التوزيع. ويستخدم اختبار حسن المطابقة لتحديد فيما إذا كان المجتمع له توزيع نظري معين، ويعتمد الاختبار على توزيع مربع " كاي " X^2 distribution والذي سبق ذكره في مباحث سابقة من هذا الكتاب. وكما تم عمله في الاختبارات السابقة، فإن هذا الاختبار يتبع الخطوات التالية:

الخطوة الأولى: تحديد الفرضية المراد اختبارها H_0 والفرضية البديلة كآلاتي:

المتغير تحت الدراسة له توزيع محدد: H_0

المتغير تحت الدراسة ليس له توزيع محدد: H_1

الخطوة الثانية: تحديد مستوى المعنوية α

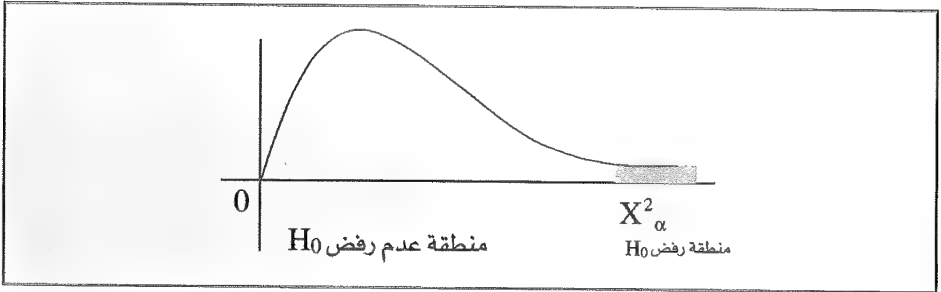
الخطوة الثالثة: حساب إحصاء الاختبار والتي يرمز لها بالرمز X^2 كآلاتي:

$$X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

حيث أن O_i و E_i هما التكرار المشاهد في العينة والتكرار المتوقع تحت التوزيع المحدد في الفرضية H_0 على التوالي.

وان $E_i = np_i$ بالاعتماد على n والذي يمثل حجم العينة و P_i الذي يمثل التكرار النسبي (الاحتمال) الذي نحصل عليه باستخدام التوزيع المحدد من الفرضية H_0 المراد اختبارها.

الخطوة الرابعة: تحديد منطقة الرفض بالاستناد إلى قيمة α واستخدام توزيع "مربع كاي" بدرجات حرية مساوية إلى $(k-1)$ حيث أن k تمثل عدد قيم المتغير تحت الدراسة بالشكل التالي:



الخطوة الخامسة: اتخاذ القرار بشأن الفرضية.

مثال (7)

إذا علمت أن توزيع الجرائم التي سجلت في إحدى المدن سنة 1995 كانت كآلاتي:

نوع الجريمة	قتل	إحتذاء	سرقاة	شغب
التكرار النسبي	0.012	0.611	0.323	0.611

وباستخدام عينة عشوائية من 500 جريمة حصلت في العام الماضي وجدنا أن:

نوع الجريمة	قتل	اعتداء	سرقة	شغب
العدد	9	26	144	321

هل أن توزيع الجريمة لهذه السنة مشابه لتوزيع الجريمة سنة 1995 . باستخدام

$$\alpha = 0.01$$

الحل

علينا لإجراء الاختبار اتباع الخطوات السابق ذكرها وتحديد الفرضية المراد اختبارها H_0 والفرضية البديلة H_1 لتكون:

H_0 : توزيع الجريمة لهذه السنة يشابه توزيع الجريمة لسنة 1995

H_1 : التوزيعان مختلفان

أما لحساب إحصاءة الاختبار X^2 ، علينا الرجوع إلى الجدول التالي:

نوع الجريمة	التكرار المشاهد O_i	التكرار المتوقع E_i	$\frac{(O_i - E_i)^2}{E_i}$
قتل	9	$(500)(0.012)=6.0$	1.500
اعتداء	26	$(500)(0.054)=27.0$	0.037
سرقة	144	$(500)(0.323)=161.5$	1.896
شغب	321	$(500)(0.611)=305.5$	0.786
	500	500	4.219

وبذلك فإن:

$$X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = 4.219$$

وبالرجوع إلى جدول X^2 بدرجة حرية $3 = 4 - 1 = k - 1$ نجد أن $X^2_{0.01} = 11.345$

وبذلك فإن منطقة الرفض تتحدد بالشكل:



وبما أن إحصاء الاختبار تقع ضمن عدم الرفض، لذلك لا نرفض H_0 ، أي أن توزيع الجريمة لهذه السنة مشابهة لتوزيع الجريمة سنة 1995.

8-8 اختبار الاستقلالية Test for Independence

اختبار الاستقلالية بين ظاهرتين أو بين متغيرين X و Y هو اختبار يعتمد على أن تكون البيانات ثنائية Binary Data بشكل ما يسمى بجداول التوافق Contingency Tables والتي سبق ذكرها عند عرض البيانات الثنائية واختبار الاستقلالية يعتبر من الاختبارات المهمة لدراسة العلاقة بين متغيرين وعن طريقه يمكن التعرف على نوعية العلاقة ووجودها المعنوي. ويعتمد هذا الاختبار أيضا على مربع كاي X^2 distribution، ويتمثل بالخطوات التالية:

الخطوة الأولى: تحديد الفرضية المراد اختبارها H_0 والفرضية البديلة H_1 لتكون:

H_0 : المتغيران تحت الدراسة مستقلان

H_1 : المتغيران تحت الدراسة غير مستقلان

الخطوة الثانية: تحديد مستوى المعنوية α .

الخطوة الثالثة: حساب إحصاء الاختبار، والتي يرمز لها بالرمز X^2 بالشكل:

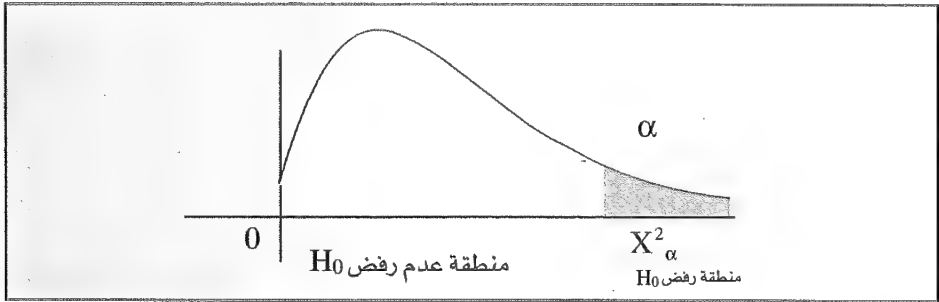
$$X^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

حيث أن O_{ij} و E_{ij} تمثلان التكرار المشاهد في العينة والتكرار المتوقع تحت فرضية الاستقلالية واللذان يخصان الصف i والعمود j على التوالي. وأن

$$E_{ij} = \frac{R_i C_j}{n}$$

بالاعتماد على R_i والذي يمثل مجموع التكرارات للصف i و C_j مجموع التكرارات للصف j ، أما n فتمثل حجم العينة.

الخطوة الرابعة: بالرجوع لتوزيع X^2 وبدرجة حرية مساوية إلى $(r - 1)(c - 1)$ حيث أن r يمثل عدد الصفوف و c يمثل عدد الأعمدة، فإن منطقة الرفض وباستخدام قيمة α المحددة تتخذ الشكل:



الخطوة الخامسة: اتخاذ القرار بشأن الفرضية.

مثال (8)

الجدول الثنائي يمثل توزيع 500 طالب حسب أوزانهم وأطوالهم. هل أن الأوزان مستقلة عن الأطوال؟ استخدم $a = 0.05$.

الأوزان \ الأطوال	50-	60-	70-80	المجموع
150 -	64	54	85	200
160 -	50	66	84	200
170 - 180	36	30	34	100
المجموع	150	150	200	500

علينا تحديد الفرضية H_0 والبديلة H_1 لتكون:

H_0 : الأوزان مستقلة عن الأطوال :

H_1 : الأوزان غير مستقلة عن الأطوال :

أما لحساب إحصاء الاختبار X^2 ، فعلينا القول بأن الجدول المعطى في السؤال يمثل التكرارات المشاهدة O_{ij} وعلينا احتساب التكرارات المتوقعة تحت فرضية الاستقلالية باستخدام:

$$E_{ij} = \frac{R_i C_j}{n}$$

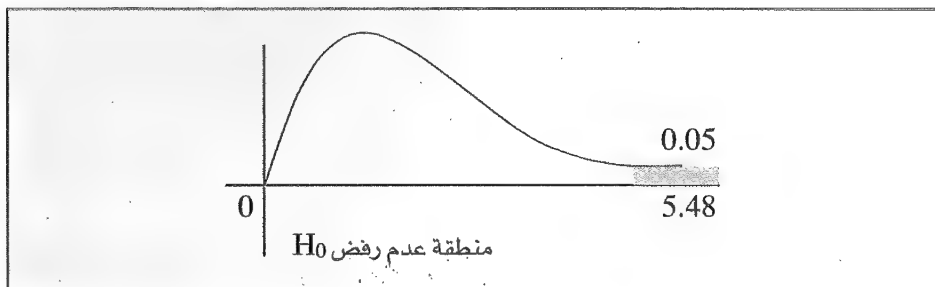
وبذلك فإن جدول التكرارات المتوقعة E_{ij} سيكون بالشكل التالي:

الأوزان الأطوال	50-	60-	70-80
150 -	$\frac{(200)(150)}{500} = 60$	$\frac{(200)(150)}{500} = 60$	$\frac{(200)(200)}{500} = 80$
160 -	$\frac{(200)(150)}{500} = 60$	$\frac{(200)(150)}{500} = 60$	$\frac{(200)(200)}{500} = 80$
-180 170	$\frac{(100)(150)}{500} = 30$	$\frac{(100)(150)}{500} = 30$	$\frac{(100)(150)}{500} = 80$

$$\begin{aligned}
 X^2 &= \sum_i \sum_j \frac{(O_i - E_i)^2}{E_{ij}} \\
 &= \frac{(64-60)^2}{60} + \frac{(54-60)^2}{60} + \frac{(82-80)^2}{80} + \frac{(50-60)^2}{60} + \frac{(66-60)^2}{60} + \\
 &\quad + \frac{(84-80)^2}{80} + \frac{(36-30)^2}{30} + \frac{(30-30)^2}{30} + \frac{(34-40)^2}{40} \\
 &= \frac{16}{60} + \frac{36}{60} + \frac{4}{80} + \frac{100}{60} + \frac{36}{60} + \frac{16}{80} + \frac{36}{30} + \frac{0}{30} + \frac{36}{40} \\
 &= 5.48
 \end{aligned}$$

وباستخدام $\alpha = 0.05$ و جدول " مربع كاي " بدرجات حرية مساوية إلى:

فإن منطقة الرفض $(r - 1) (c - 1) = (3 - 1) (3 - 1) = 4$



وبما أن إحصاءة الاختبار تقع ضمن منطقة عدم الرفض، لذلك لا نرفض H_0 ، أي أن المتغيران الأوزان والأطوال مستقلان .

لتطبيقات SPSS:

يمكن إيجاد مربع كاي (X^2) لإيجاد هل هناك اقتران بين المتغيران أم أنهما مستقلان نقوم بالخطوات التالية:

نختار الخيار Analyze من القائمة الرئيسية إلى SPSS ومن هذه القائمة نختار الخيار Descriptive ومن هذه القائمة نختار الخيار Crosstabs وهنا نحدد المتغيرات في هذه الشاشة فنضع رتب تكرارات المتغير المعتمد في حقل Row (s) ونضع رتب تكرارات المتغير المستقل في حقل Column (s) وبعد ذلك ننقر على خيار Statistics في هذه الشاشة ومن هذه الشاشة نختار الخيارات Chi-square و correlations وننقر على الزر continue وبعد ذلك ننقر على الزر ok للتنفيذ.

أسئلة

الفصل الثامن

1- معمل للورق يشترط أن يكون معدل طول الأشجار الصالحة للقطع هي 40 قدم فاكتر، سحبت عينة عشوائية من 7 أشجار من غابة اختيرت لهذا الغرض وكانت أطوالها كالآتي:

40.1	39	42	41.7	42.2	41.5	41
------	----	----	------	------	------	----

اختبر فيما إذا كان باستطاعة الشركة القول بأن معدل أطوال الأشجار في تلك الغابة هي 40 قدم. استخدم $\alpha = 0.05$.

2- سحبت عينة عشوائية أولى بالحجم 20 وعينة عشوائية ثانية بالحجم 15 من مجتمعين وكانت القياسات التي تم الحصول عليها هي:

(I) المجتمع				(II) المجتمع		
9.0	15.6	25.6	31.3	10.1	11.1	13.5
21.1	26.9	24.6	20.0	12.0	18.2	10.3
24.8	16.5	26.0	25.1	9.2	7.0	14.2
17.2	50.1	18.7	26.1	15.8	13.6	13.2
18.2	25.4	22.0	23.3	8.8	12.5	21.5

أ) اختبر الفرضية التي تنص على أن التوزيع للمجتمع (I) يختلف عن التوزيع للمجتمع (II)، استخدم $\alpha = 0.05$.

ب) اختبر الفرضية التي تنص على أن وسط المجتمع (II) اكبر من وسط المجتمع (I) مستخدماً $\alpha = 0.01$.

3- أحد الاقتصاديين يرغب في اختبار الفرق بين المستوى الاقتصادي لثلاث مناطق مختلفة عن طريق سحب عينات مختلفة من المناطق الثلاث وكانت البيانات كالآتي:

الطريقة الأولى	4.7	5.8	3.8	5.6	6.2	5.2
الطريقة الثانية	7.6	4.9	5.2	6.8	6.7	5.9
الطريقة الثالثة	4.3	4.2	6.2	3.9	4.8	5.1

هل أن البيانات تعطي الأدلة الكافية على وجود اختلافات. استخدم $\alpha = 0.01$.

4- نتائج استخدام ثلاث طرق مختلفة للتدريس أعطت النتائج التالية:

الطريقة الأولى	8.3	8	7.5	7	6.5
الطريقة الثانية	9.0	8	8	9.5	
الطريقة الثالثة	7.6	7	7.3	7.5	

هل أن البيانات تشير إلى وجود اختلاف. استخدم $\alpha = 0.05$.

5- اختبر فرضية الاستقلالية بين التصنيفين A و B باستخدام أصول التوافق

التالي، استخدم هل أن البيانات تعطي الأدلة الكافية على وجود اختلافات.

استخدم $\alpha = 0.05$.

	B1	B2	B3	B4
A1	22	38	29	51
A2	42	27	68	53
A3	26	85	102	68

6- هل أن نوع الحادث مستقل عن عمر السائق. اختبر مستخدماً $\alpha = 0.01$.

العمر	نوع الحادث		
	I	II	III
أقل من 25	9	17	5
25 فأكثر	61	13	12

7- البيانات التالية تمثل متغيرين X و Y كالآتي:

X:	1-	2	5-	1-	3	0	4
Y:	1-	1	4-	1	3	1	2

أوجد قيمة معامل الارتباط r ، ثم اختبر فيما إذا كان الارتباط موجود بين المتغيرين X و Y مستخدماً $\alpha = 0.05$.

8- البيانات التالية تمثل عدد ساعات الدراسة والدرجات لمجموعة من التلاميذ، اختبر معنوية الارتباط مستخدماً $\alpha = 0.01$.

عدد ساعات الدراسة :	0	1	2	3	4	5
الدرجة :	70	81	70	65	85	83

9

الفصل التاسع

الأرقام القياسية والسلاسل الزمنية

Index Numbers and Time Series

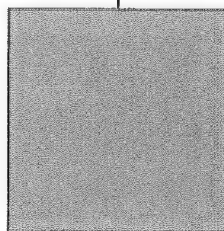
9-1 مقدمة

9-2 السلسلة الزمنية والأرقام القياسية

9-3 تحليل السلاسل الزمنية

الفصل الأسع

9



الفصل التاسع

الأرقام القياسية والسلاسل الزمنية

Index Numbers and Time Series

9-1 مقدمة Introduction

سنقوم في هذا الفصل بدراسة قياسات أو مشاهدات مأخوذة لعدد من الفترات الزمنية والتي يطلق عليها اسم السلاسل الزمنية Time Series عن طريق وصف هذه السلاسل وتحليلها.

وسيتكون هذا الفصل من عرض لبعض أنواع الأرقام القياسية ووصف السلسلة الزمنية حسب مكوناتها أو مركباتها ومن ثم تحليل هذه السلاسل باستخدام بعض من الطرق الإحصائية المناسبة ومن ثم استخدام تلك الطرق لأجل تقدير قيم السلاسل الزمنية في فترات مستقبلية، أي القيام بعملية التنبؤ. وسيتم كل ذلك من خلال ما يلي:

9-2 السلسلة الزمنية والأرقام القياسية Time Series : Index Numbers

يمكن تعريف السلسلة الزمنية Time Series على أنها مجموعة من القياسات أو المشاهدات أو البيانات والمرتبة حسب فترات زمنية متعددة. ويفضل لقراءة هذه السلاسل استخدام عدد مناسب وليس قليلا من تلك الفترات، حيث أن التغيرات والتأثيرات يمكن أن تظهر وبشكل واضح لسلسلة زمنية بعدد من الفترات ولتكن 03 فترة افضل من العدد 10 فترة.

والسلاسل الزمنية عادة ما تكون قياساتها الأسعار، فمثلا نتكلم عن أسعار الذهب أو أسعار النفط أو غير ذلك من السلع الإنتاجية والاستهلاكية والاستراتيجية وغيرها، كما ويمكن أن تكون القياسات هي الكميات، ولكن ما يميز السلسلة الزمنية عن غيرها من أنواع البيانات هو الترتيب Order، أي أن هناك ترتيب معين للفترات

الزمنية فنقول أن السلسلة تمثل الأسعار للفترات 1, 2, 3, ..., n, أو نقول بأن الأسعار للسنوات 1960، 1961، ... وهكذا.

أما عن وصف هذه السلاسل الزمنية فيمكن وبشكل مبسط حساب التغيرات التي تحدث في هذه القياسات لفترات قصيرة عن طريق استخدام الأرقام القياسية. وهنا نقول بأن لدينا:

9-2-1 الرقم القياسي البسيط Simple Index Number:

للسلسلة الزمنية المحددة بالفترات المختلفة نختار إحدى القيم لتكون ما يسمى بسنة الأساس base period ومن ثم فإن السنوات الأخرى تمثل سنوات مقارنة. الرقم القياسي في السنة t سيكون

$$I_t = \frac{P_t}{P_0} (100\%)$$

حيث أن P_0 هو القياس أو السعر في سنة الأساس.
أما P_t فهو القياس أو السعر في سنة المقارنة.

مثال (1)

البيانات التالية تمثل أسعار الذهب (دولار/ أونس) للسنوات 1970 إلى 1985

السنة	1970	1971	1972	1973	1974	1975	1976	1977	1978
سعر الذهب	36	41	59	98	160	161	125	148	194

1979	1980	1981	1982	1983	1984	1985
308	613	460	376	424	361	318

احسب الأرقام القياسية وفسر معناها.

باستخدام سنة 1976 كسنة أساس يمكن استخراج الأرقام القياسية التالية كما تظهر في جدول (1) التالي:

جدول (1)

الأرقام القياسية لأسعار الذهب

السنة	سعر الذهب	الأرقام القياسية
1970	36	28.8%
1971	41	32.8%
1972	59	47.2%
1973	98	78.4%
1974	160	128.0%
1975	161	128.8%
1976	125	100.0%
1977	148	118.4%
1978	194	155.2%
1979	308	246.4%
1980	613	490.4%
1981	460	368.0%
1982	376	300.8%
1983	424	339.2%
1984	361	288.8%
1985	318	254.4%

ويلاحظ من قيم الأرقام القياسية أن هذه الأرقام يمكن أن تكون أكبر من 100 أو أن تكون أقل من 100، حيث أن الرقم القياسي لسنة الأساس هو دائما 100 وذلك لأنه يقارن السعر لتلك السنة مع تلك السنة الأساس.

الرقم القياسي سنة 1982 هو 300.8 أي أن الأسعار تقريبا ثلاثة أمثال الأسعار في

سنة الأساس 1976، أما الرقم القياسي لسنة 1973 فهو 78.4 أي أن الأسعار تقريبا ثلاثة أرباع الأسعار في سنة الأساس 1976 وهكذا يمكن تفسير هذه الأرقام القياسية والتي تظهر الاختلافات والتغيرات التي تظهر في قيم السلسلة على الفترات المختلفة.

2-2-9 الرقم القياسي التجميعي البسيط Composite Index Numbers:

إذا كانت السلسلة الزمنية تمثل أسعار أكثر من سلعة لفترات الزمنية المختلفة فيمكن عندئذ حساب الرقم القياسي التجميعي وذلك بتجميع الأسعار للسلع المختلفة ولكل فترة زمنية ومن ثم اختيار إحدى الفترات لتكون سنة الأساس ومن ثم تنسب كل مجموع على ذلك المجموع لسنة الأساس لتكوين الأرقام القياسية التجميعية البسيطة والتي تظهر بالشكل:

$$I_t = \frac{\sum_{j=1}^k P_{tj}}{\sum_{j=1}^k P_{0j}} (100\%)$$

حيث أن $\sum P_{0j}$ يمثل مجموع أسعار سلع مختلفة في سنة الأساس.

وأن $\sum P_{tj}$ يمثل مجموع أسعار السلع المختلفة في سنة المقارنة t.

أما k فتمثل عدد السلع.

مثال (2)

البيانات التالية تمثل الإنفاق على الغذاء، الملابس، الإيجار وغيرها لأسرة معينة خلال سنتين:

السنة الأولى :	7	3	2	3
السنة الثانية :	29	12	7	7

أوجد الرقم القياسي لكل بند من بنود الإنفاق لتلك الأسرة. وأوجد الرقم القياسي التجميعي لإنفاق تلك الأسرة مستخدما السنة الأولى كسنة أساس.

الأرقام القياسية لكل بند من بنود الإنفاق وعلى التوالي هي:
 233,333 % 350,000 % 400,000 % 414,286 % : الأرقام القياسية

أما الرقم القياسي التجميعي فهو:

$$I_t = \frac{29+12+7+7}{7+3+2+3} (100\%) = 366667\%$$

9-2-3 الرقم القياسي لاسبير Laspeyres Index Number:

ويعتبر من الأرقام القياسية المرجحة، بمعنى أن هناك أوزان معينة لكل سعر في السلسلة الزمنية. وبذلك فإن الرقم القياسي لاسبير (حسب اسم العالم لاسبير) ويتم حسابه بقسمة مجموع مرجح لسنة مقارنة على مجموع مرجح لسنة أساس. ويكون الترجيح لأسعار سنوات الأساس والمقارنة بأوزان سنة الأساس. وبذلك فإن هناك رقمان قياسييان هما:

$$I_{AL} = \frac{\sum_{j=1}^k P_{tj} q_{0j}}{\sum_{j=1}^k P_{0j} q_{0j}} (100\%) \quad \text{الرقم القياسي لاسبير التجميعي}$$

$$I_{RL} = \sum \frac{P_{tj}}{P_{0j}} q_{0j} (100\%) \quad \text{والرقم القياسي لاسبير النسبي}$$

9-2-4 الرقم القياسي لباش Passche Index Number:

أما الرقم القياسي المرجح الآخر فهو الرقم القياسي لباش (حسب اسم العالم باش). ويتم حسابه بقسمة مجموع مرجح لسنة مقارنة على مجموع مرجح لسنة الأساس. ويكون الترجيح لأسعار سنوات الأساس والمقارنة بأوزان سنة المقارنة. وبذلك فإن هناك رقمان قياسييان هما:

$$I_{AP} = \frac{\sum_{j=1}^k p_{ij} q_{ij}}{\sum_{j=1}^k p_{0j} q_{ij}} (100\%)$$

الرقم القياسي لباش التجميعي

$$I_{RP} = \sum_{j=1}^k \frac{p_{ij}}{p_{0j}} q_{ij} (100\%)$$

الرقم القياسي لباش النسبي

9-2-5 الرقم القياسي الأمثل لفشر Fisher's Ideal Index Number:

وهذا الرقم القياسي يعتبر من افضل الأرقام القياسية على الإطلاق. ويعتمد على الرقمين القياسيين إلى لاسبير وباش في نفس الوقت، وهو أيضا نوعان تجميعي ونسبي وبالشكل التالي:

$$I_{A.F.I.I} = \sqrt{(I_{AL})(I_{AP})}$$

الرقم القياسي الأمثل لفشر التجميعي

$$I_{R.F.I.I} = \sqrt{(I_{RL})(I_{RP})}$$

أما الرقم القياسي الأمثل لفشر النسبي فهو

مثال (3)

البيانات التالية تمثل أسعار وأوزان ثلاث سلع (رز، سكر وكاز) للسنتين 1965 ، 1975. أوجد الرقم القياسي للاسبير وباش وفشر التجميعية والنسبية، باستخدام 1965 كسنة أساس.

السلعة	السعر		الأوزان	
	1965	1975	1965	1975
رز	150	230	0.53	0.60
سكر	10	46	0.17	0.20
كاز	30	34	0.30	0.20
			1.00	1.00

لإيجاد الرقم القياسي للاستخدام أوزان سنة الأساس 1965 للترجيح وبذلك فإن:

$$I_{AL} = \frac{(230)(0.53) + (46)(0.17) + (34)(0.30)}{(150)(0.53) + (10)(0.17) + (30)(0.30)} (100\%) = 155\% \text{ (التجميعي)}$$

$$I_{RL} = \left[\frac{230}{150} (0.53) + \frac{46}{10} (0.17) + \frac{34}{30} (0.30) \right] (100\%) = 193.5\% \text{ (النسبي)}$$

أما الرقم القياسي لباش فنستخدم أوزان سنة المقارنة 1975 للترجيح وبذلك فإن:

$$I_{AP} = \frac{(230)(0.60) + (46)(0.20) + (34)(0.20)}{(150)(0.60) + (10)(0.20) + (30)(0.20)} (100\%) = 157\% \text{ (التجميعي)}$$

$$I_{RP} = \left[\frac{230}{150} (0.60) + \frac{46}{10} (0.20) + \frac{34}{30} (0.20) \right] (100\%) = 207\% \text{ (النسبي)}$$

أما الرقم القياسي الأمثل لفشر فسيكون:

$$I_{A.F.I.I} = \sqrt{(155)(157)} = 155.997\% \text{ (التجميعي)}$$

$$I_{R.F.I.I} = \sqrt{(193)(207)} = 199.877\% \text{ (النسبي)}$$

9-2-6 الرقم القياسي الحقيقي للدخل The True Income Index Number

يعرف هذا الرقم القياسي والذي يرمز له بالرمز I_T بالشكل التالي:

$$I_T = \frac{\text{الرقم القياسي للدخل}}{\text{الرقم القياسي للتكاليف}} = \frac{I_I}{T_C}$$

حيث:

I_T تمثل الرقم القياسي الحقيقي للدخل.

I_I تمثل الرقم القياسي للدخل:
 I_C تمثل الرقم القياسي للتكاليف.

مثال (4)

البيانات في الجدول التالي تمثل أسعار وكميات أربع سلع خلال سنين متتالية للفرد:

السلع				
A	B	C	D	أسعار السنة الأولى 2000 (الأساس)
10	15	20	25	أسعار السنة الثانية 2001 (المقارنة)
20	25	30	35	
20	25	15	30	الكميات لسنة 2000 (الأساس)
25	30	20	35	الكميات لسنة الأولى 2001 (المقارنة)

وبافتراض أن هذه الكميات الأربعة تمثل تكاليف المعيشة للفرد للسنة وأن الدخل الشهري للفرد لسنة 2000 كانت 400 دينار ولسنة 2001 كان الدخل 500 دينار. أوجد الرقم القياسي للدخل الحقيقي للفرد وفسر النتائج.

الحل

$$I_I = \frac{500}{400} = 1.25 \quad \text{الرقم القياسي للدخل:}$$

$$I_C = \frac{110}{90} = 1.222 \quad \text{والرقم القياسي لتكاليف المعيشة:}$$

$$I_T = \frac{1.25}{1.222} = 102.3\% \quad \text{وبذلك فإن الرقم القياسي الحقيقي للدخل:}$$

ويفسر بوجود رفاهية وبنسبة قليلة جداً 2.3%.

9-3 تحليل السلاسل الزمنية:

سيتم تحليل السلاسل الزمنية عن طريق دراسة مكوناتها أولاً ومن ثم استخدام بعض من الطرق المتوفرة لتحليلها واستخداماتها لأغراض التنبؤ Forecasting.

يمكن كتابة النموذج الخاص بالسلسلة الزمنية عن طريق عرض التأثير الموجود فيها إلى أكثر من مركب، وسنؤكد هنا على استخدام الشكل الأكثر استخداماً والذي يدعى بنموذج الجمع Additive Model والذي يأخذ الصورة العامة التالية:

$$y_t = T_t + C_t + S_t + R_t$$

حيث أن y_t يمثل قيمة السلسلة الزمنية عند الفترة t وأن T_t يمثل ما يسمى بالاتجاه العام Trend للسلسلة، وقد يكون الاتجاه العام موجب (طردي) أو سالب (عكسي).

أما C_t فيمثل التأثير الدوري Cyclical effect والذي يوصف التغيرات في السلسلة حول اتجاهها العام، والذي يرجع بصورة عامة للظروف الاقتصادية والإدارية كما أن S_t يمثل التأثير الموسمي Seasonal effect والذي يوصف كما يدل الاسم على ذلك التأثير الذي يظهر في السلسلة بشكل موسمي ضمن فترة معينة، كتأثير بيع بضاعة معينة في موسم معين.

أما R_t فيمثل تأثير الخطأ Residual effect وهو التأثير المتبقي بعد إزالة جميع التأثيرات السابق ذكرها من السلسلة. وكذلك يمكن أن يوصف جميع الظروف غير الطبيعية والتي قد تحدث وتؤثر على قيم السلسلة.

أما عن بعض الطرق المستخدمة والبسيطة لتحليل السلاسل الزمنية والتي سيتم التطرق إليها هنا فهي:

9-3-1 طريقة الانحدار الخطي البسيط Simple Linear Regression:

لقد تم وصف هذه الطريقة وبشكل وافي في فصل سابق من هذا الكتاب، ولن يتم تكرار ذلك الوصف سوى انه سيتم تحليل سلسلة زمنية بهذه الطريقة والتي تعتبر من السبل المناسبة لتحديد الاتجاه العام للسلسلة.

مثال (5)

بالرجوع لبيانات مثال (1) السابق. أوجد خط الاتجاه العام للسلسلة الزمنية واستخدمها لتقدير سعر الذهب سنة 1990.

الحل

بالرجوع لبيانات المثال وبافتراض أن المتغير المستقل t والذي يمثل الفترات الزمنية وافترض أن المتغير التابع y هو سعر الذهب فإن تقدير خط الاتجاه العام أو تقدير خط الانحدار هو:

$$\hat{y}_t = \beta_0 + \beta_1 t$$

$$\hat{y}_t = -5.650 + 29.209t$$

وبذلك فإن تقدير سعر الذهب سنة 190 والذي يمثل $t = 21$ فهو

$$\hat{y}_t = -5.650 + 29.209(21) = 607.739$$

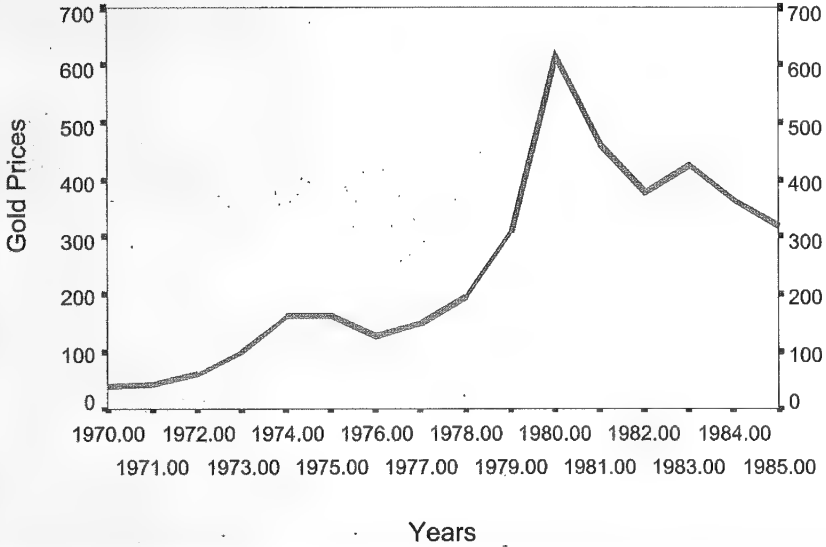
أما عند رسم نقاط الانتشار وخط الاتجاه العام فيظهر في شكل (1) أدناه:

تطبيقات SPSS:

ويمكن إيجاد المعادلة باستخدام الانحدار الخطي البسيط في البرنامج SPSS وكما يلي:

نختار Analyze من القائمة الرئيسية إلى SPSS ثم نختار Regression من هذه القائمة ونختار Linear من هذه القائمة ثم نضع متغير القيم في حقل dependent المتغير المعتمد. ونضع متغير متسلسل التاريخ (سنين أو أشهر Time) في حقل Independent وننقر على ok للتنفيذ.

Graph no.1
Original Data



9-3-2 طريقة المتوسطات المتحركة Moving Averages:

وهي طريقة مناسبة لإزالة التأثيرات التي تتغير بسرعة كبيرة وذلك في محاولة تقليل تأثير مثل هذه التغيرات ولكشف تأثير الاتجاه العام لتلك السلسلة.

المتوسط المتحرك لـ N من النقاط N -Points Moving Average لسلسلة زمنية هو المتوسط لقيم السلسلة الزمنية لـ N من الفترات الزمنية المقاربة، فمثلاً لإيجاد المتوسط المتحرك لـ 3 قيم عند الفترة الزمنية t فهو:

$$M_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}$$

حيث أن: y_t يمثل قيمة السلسلة عند الفترة t .

y_{t-1} يمثل قيمة السلسلة عند الفترة $t-1$ (القيمة السابقة).

y_{t+1} يمثل قيمة السلسلة عند الفترة $t+1$ (القيمة اللاحقة).

وهكذا لبقية قياسات المتوسطات المتحركة.

اختيار العدد N لإيجاد المتوسطات المتحركة مهم جداً، حيث أن القيم الكبيرة إلى N يوصل إلى الحصول على سلسلة أكثر نعومة ولكن ذلك قد يؤثر على فقدان بعض القيم للسلسلة الزمنية، كما سنرى ذلك واضحاً من خلال الأمثلة التالية والتي سيتم حساب أكثر من شكل من أشكال المتوسطات المتحركة فيها.

استخدم بيانات المثال السابق والتي تمثل أسعار الذهب بالدولار/ الأونس للسنوات 1970 إلى 1985. المطلوب إيجاد المتوسطات المتحركة باستخدام 3 فترات زمنية، ثم 5 فترات زمنية وارسم النتائج.

مثال
(6)

النتائج تظهر في جدول (2) التالي:

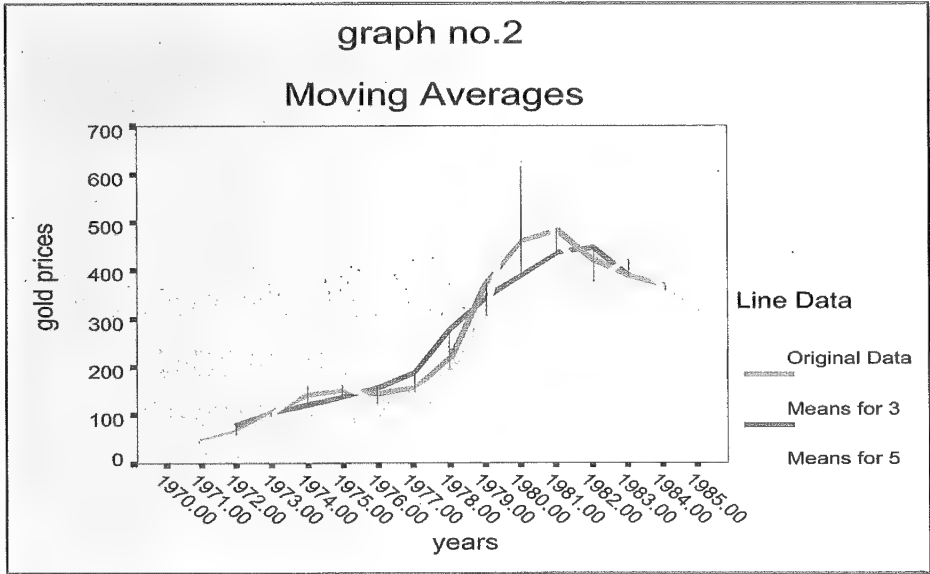
الجدول

جدول (2)

المتوسطات المتحركة لأسعار الذهب

السنة	سعر الذهب	متوسطات متحركة لـ 3 فترات	متوسطات متحركة لـ 5 سنوات
1970	36	-	-
1971	41	45.33	-
1972	59	66.00	78.80
1973	98	105.67	103.80
1974	160	139.67	120.60
1975	161	148.67	138.40
1976	125	144.67	157.60
1977	148	155.67	187.20
1978	194	216.67	277.60
1979	308	371.67	344.60
1980	613	460.33	390.20
1981	460	483.00	436.20
1982	376	420.00	446.80
1983	424	387.00	387.80
1984	361	367.67	-
1985	318	-	-

أما رسم هذه المتوسطات مع السلسلة الزمنية فيظهر في شكل (2) أدناه:



ويلاحظ بأن السلسلة ذات الأوساط المتحركة باستخدام 5 نقاط أكثر انسيابية أو نعومة More Smoother من السلسلة ذات 3 نقاط.

أما عند استخدام عدد زوجي من النقاط لاحتساب المتوسطات المتحركة فيوصلنا لأوساط متحركة غير متمركزة، أي لا تقع عند إحدى الفترات الموجودة بل تقع بين فترتين، ولذلك علينا إعادة تمركزها بإيجاد متوسطات جديدة لهذه المتوسطات المتحركة وسيتم توضيح ذلك في المثال التالي:

مثال (7)

استخدم بيانات المثال السابق لإيجاد متوسطات متحركة باستخدام 4 فترات زمنية ثم ارسم النتائج.

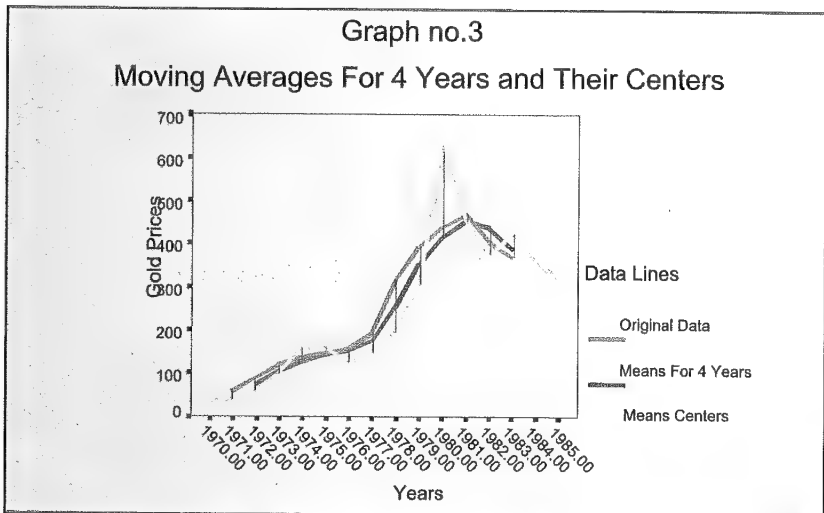
الحل

النتائج تظهر في جدول (3) التالي:

جدول (3)
المتوسطات المتحركة لأسعار الذهب

السنة	سعر الذهب	متوسطات متحركة (شهر متمركزة) 4 فترات	متوسطات متحركة متمركزة
1970	36		-
1971	41		-
		58.50	
1972	59		74.00
		89.50	
1973	98		104.375
1974	160	119.25	127.625
1975	161	136.00	142.250
1976	125	148.50	152.750
1977	148	157.00	175.375
1978	194	193.75	254.75
1979	308	315.75	354.75
1980	613	393.75	416.50
1981	460	439.25	453.75
1982	376	468.25	436.75
1983	424	405.25	387.50
1984	361	369.75	-
1985	318	-	-

أما عن رسم المتوسطات أعلاه والسنوات فيظهر في شكل (3) أدناه:



9-3-3 طريقة التنعيم Exponential Smoothing:

من مساوئ طريقة المتوسطات المتحركة السابق ذكرها هو فقدان بعض قيم السلسلة وخصوصاً عند أطرافها كما رأينا ذلك واضحاً من خلال الأمثلة، وكذلك استخدام تلك الطريقة لقيم مستقبلية في السلسلة لحساب المتوسطات. ولذلك فإن طريقة التنعيم والتي سيتم الحديث عنها الآن يمكن تفضيلها عن الطريقة السابقة وذلك لأنها لا تفقد أي من قيم السلسلة، إضافة لاستخدام طريقة التنعيم كطريقة مناسبة للتنبؤ بالقيم المستقبلية Forecasting وبذلك فإنها طريقة مناسبة لتحليل السلاسل الزمنية واستخداماتها للتنبؤ.

وطريقة التنعيم تعتمد على إعطاء أوزان موجبة Positive Weights لقيمة سابقة وقيمة حالية فقط. حيث أن ω هو الوزن أو الثابت المستخدم في هذه الطريقة بحيث أن ω تقع في الفترة بين الصفر والواحد. ومن ثم إيجاد القيم المتوقعة E_t في الفترة t باستخدام العلاقات التالية:

$$E_1 = y_1$$

$$E_2 = \omega y_2 + (1-\omega) E_1$$

$$E_3 = \omega y_3 + (1-\omega) E_2$$

.

.

$$E_t = \omega y_t + (1-\omega) E_{t-1}$$

وبذلك فإن هذه الطريقة تعطي الوزن ω للقيمة الحالية والوزن $(1-\omega)$ للقيمة السابقة والتي تم احتسابها باستخدام هذه الطريقة. وسيتم توضيح حسابات هذه الطريقة في مثال (7) اللاحق.

أما عن استخدام هذه الطريقة للتنبؤ فسيتم باستخدام آخر قيمة محسوبة وهي E_t للتنبؤ بالقيمة المستقبلية الجديدة ولتكن F_{t+1} ، أي أن

$$F_{t+1} = E_t$$

وبافتراض أن جميع التأثيرات غير موجودة فإن تقدير أي قيمة مستقبلية ستكون هي نفس القيمة الأخيرة التي تم احتسابها وبذلك فإن

$$F_{t+2} = E_t$$

$$F_{t+3} = E_t$$

مثال (8)

مستخدما نفس بيانات المثال السابق. أوجد القيم المتوقعة بطريقة التنعيم وارسم النتائج مستخدما $\omega = 0.3$ ثم $\omega = 0.7$.

الحل

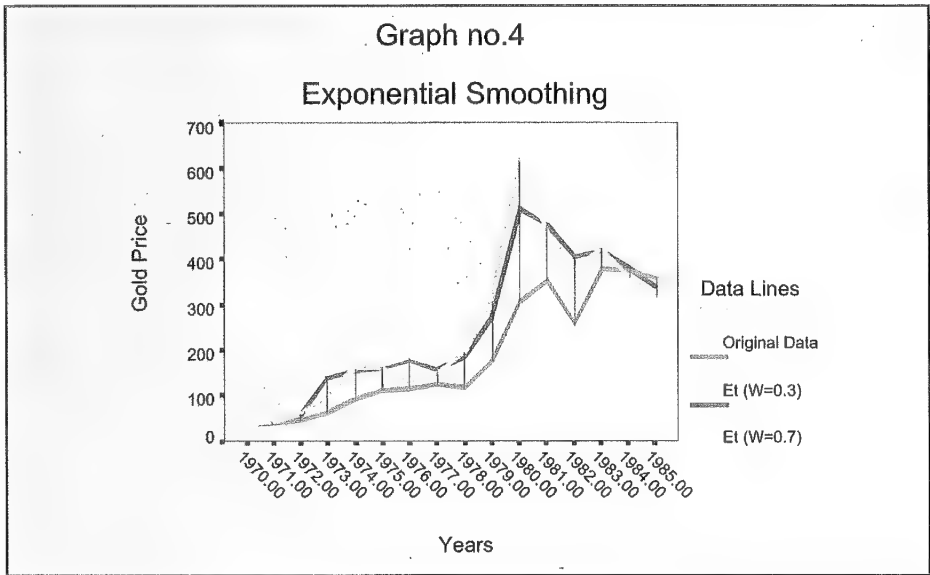
النتائج تظهر في جدول (4) التالي:

جدول (4)

استخدام طريقة التنعيم لأسعار الذهب

السنة	سعر الذهب (yt)	$E_t (\omega = 0.3)$	$E_t (\omega = 0.7)$
1970	36	36.000	36.000
1971	41	37.500	39.500
1972	59	43.950	53.150
1973	98	60.165	137.200
1974	160	90.115	153.160
1975	161	111.381	158.648
1976	125	115.467	175.000
1977	148	125.227	156.100
1978	194	116.400	182.630
1979	308	173.880	270.389
1980	613	305.616	510.217
1981	460	351.931	475.065
1982	376	259.152	405.719
1983	424	378.606	418.516
1984	361	373.324	378.255
1985	318	356.727	336.076

أما رسم قيم السلسلة الزمنية والقيم المتوقعة للحالتين أعلاه فتظهر في شكل (4) أدناه:



تطبيقات EXCEL:

يمكن عمل هذا الاختبار بواسطة EXCEL كما يلي:

EXCEL - Tools - data analysis - exponential smoothing. The dampind factor is (1-w).

أسئلة

الفصل التاسع

1- البيانات التالية تمثل أسعار النفط الخام للسنوات 1970 إلى 1985 كالآتي:

السنة	1970	1971	1972	1973	1974	1975	1976	1977	1978
سعر النفط الخام (دولار/برميل)	1.80	2.18	2.48	5.18	10.46	11.51	11.51	12.70	15.40
	1979	1980	1981	1982	1983	1984	1985		
	18.00	28.00	32.00	34.00	30.00	26.00	26.00		

المطلوب:

- إيجاد الأرقام القياسية البسيطة باستخدام 1972 كسنة أساس.
 - تحليل الاتجاه العام للسلسلة الزمنية وتقدير سعر النفط الخام سنة 1986.
 - استخدام طريقة التنعيم لتقدير سعر النفط الخام سنة 1986، ستخدما $\omega=0.5$.
- 2- البيانات التالية تمثل الأسعار والأوزان للسنوات 1984، 1985، و1986 لثلاثة من الشركات الكبرى

شركة (1)			شركة (2)		شركة (3)	
	السعر	الوزن	السعر	الوزن	السعر	الوزن
1984	29.875	229.7	44.000	4843.20	114.125	27215.2
1985	22.875	487.4	43.375	11869.0	136.375	36521.8
1986	24.500	167.3	64.250	7772.9	151.500	31936.9

المطلوب:

- حساب الرقم القياسي البسيط.

- حساب الرقم القياسي التجميعي.
 - حساب الرقم القياسي للاسبير، التجميعي والنسبي.
 - حساب الرقم القياسي لباش، التجميعي والنسبي.
 - حساب الرقم القياسي الأمثل لفشر، التجميعي والنسبي.
- 3- البيانات التالية تمثل مبيعات شركة كبرى للسيارات للسنوات 1970 إلى 1985:

السنة	1970	1971	1972	1973	1974	1975	1976	1977	1978
المبيعات (مليون)	5.308	7.779	7.791	8.648	6.690	6.629	8.568	9.068	9.482
السنة	1979	1980	1981	1982	1983	1984	1985		
	8.993	7.101	6.762	6.244	7.769	8.256	9.305		

المطلوب:

- إيجاد المتوسطات المتحركة باستخدام ثلاث فترات زمنية.
- إيجاد المتوسطات المتحركة باستخدام سبع فترات زمنية.
- رسم النتائج (أ) و (ب) في رسم واحد.
- مقارنة نتائج (أ) و (ب).
- إيجاد المتوسطات المتحركة باستخدام أربع فترات زمنية.

4- افترض بيانات السلسلة الزمنية التالية:

السنة	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
المبيعات (مليون)	10919.2	11206.1	10857.0	12416.8	16234.8	22255.3	20633.0	16739.2	14852.0	15139.9

المطلوب:

- حساب الأرقام القياسية البسيطة مستخدماً سنة 1975 كسنة أساس.
 - حساب الأرقام القياسية البسيطة مستخدماً سنة 1980 كسنة أساس.
- 5- مستخدماً May 1984 كسنة أساس. أحسب وارسم الأرقام القياسية المركبة البسيطة لكل سلعة مما يلي:

(1984) : Jan. Feb. Mar. Apr May June July Aug. Sept. Oct. Nov. Dec.

سعر البيض : 96.1 92.9 79.4 91.4 68.9 61.0 59.9 58.6 58.4 55.3 61.3 58.4

سعر القهوة : 1.34 1.34 1.34 1.28 1.35 1.28 1.22 1.17 1.21 1.25 1.26 1.13

سعر الكازولين : 113.1 112.5 112.5 114.5 115.4 114.7 112.9 111.6 112.0 112.7 112.4 110.9

6- افترض بيانات السلسلة الزمنية التالية:

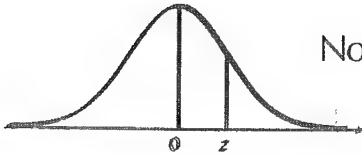
السنة :	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
المبيعات	5.308	7.779	7.791	8.684	6.690	6.629	8.568	9.068	9.482	8.993	7.101
	1981	1982	1983	1984	1985						
	6.762	6.944	7.77	8.256	9.305						

المطلوب:

- إيجاد المتوسطات المتحركة ورسمها باستخدام ثلاث سنوات.
- إيجاد المتوسطات المتحركة ورسمها باستخدام خمس سنوات.

ملحق

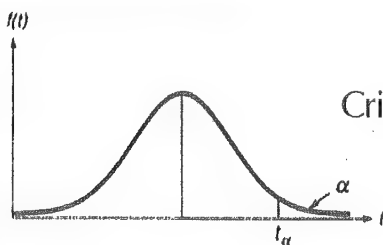
ملحق
بالجداول
الإحصائية



Normal Curve Areas

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

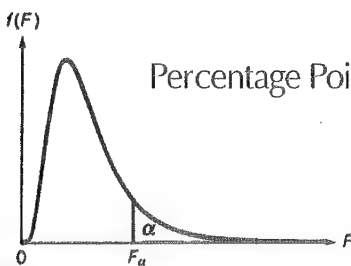
Source: Abridged from Table I of A. Hald, *Statistical Tables and Formulas* (New York: Wiley), 1952. Reproduced by permission of A. Hald and the publisher, John Wiley & Sons, Inc.



Critical Values of t

ν	t_{100}	t_{050}	t_{025}	t_{010}	t_{005}	t_{001}	t_{005}
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	1.289	1.658	1.980	2.358	2.617	3.160	3.373
120	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Source: This table is reproduced with the kind permission of the Trustees of Biometrika from E. S. Pearson and H. O. Hartley (eds.), *The Biometrika Tables for Statisticians*, Vol. 1, 3d ed., Biometrika, 1966.



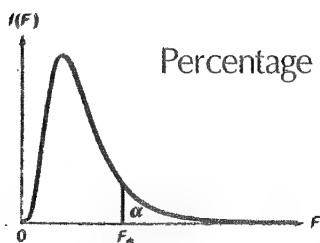
Percentage Points of the F Distribution, $\alpha = .10$

ν_2	ν_1	NUMERATOR DEGREES OF FREEDOM								
		1	2	3	4	5	6	7	8	9
1	1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
2	1	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	1	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
4	1	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
5	1	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	1	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	1	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	1	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	1	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	1	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
11	1	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
12	1	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	1	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	1	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	1	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	1	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	1	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	1	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	1	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	1	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	1	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	1	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	1	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	1	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	1	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	1	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	1	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	1	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
29	1	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	1	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
40	1	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
60	1	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
120	1	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
∞	1	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63

Source: From M. Merrington and C. M. Thompson, "Tables of Percentage Points of the Inverted Beta (F)-Distribution," *Biometrika*, 1943, 33, 73-88. Reproduced by permission of the *Biometrika* Trustees.

V2	V1	NUMERATOR DEGREES OF FREEDOM									
		10	12	15	20	24	30	40	60	120	∞
1		60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2		9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3		5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4		3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5		3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6		2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7		2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8		2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9		2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10		2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11		2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12		2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13		2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14		2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15		2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16		2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17		2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18		1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19		1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20		1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21		1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22		1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23		1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24		1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25		1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26		1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27		1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28		1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29		1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30		1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40		1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60		1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120		1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞		1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

DENOMINATOR DEGREES OF FREEDOM



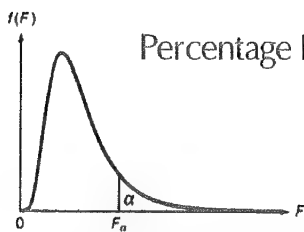
Percentage Points of the F Distribution, $\alpha = .05$

		NUMERATOR DEGREES OF FREEDOM							
	ν_1								
	ν_2	1	2	3	4	5	6	7	8
	1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9
	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37
	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85
	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45
	21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42
	22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40
	23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37
	24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36
	25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34
	26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32
	27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31
	28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29
	29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28
	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27
	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18
	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10
	120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02
	∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94

Source: From M. Merrington and C. M. Thompson, "Tables of Percentage Points of the Inverted Beta (F)-Distribution," *Biometrika*, 1 73-88. Reproduced by permission of the Biometrika Trustees.

V ₂	V ₁	NUMERATOR DEGREES OF FREEDOM									
		10	12	15	20	24	30	40	60	120	∞
1		241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2		19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3		8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4		5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5		4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6		4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7		3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8		3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9		3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10		2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11		2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12		2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13		2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14		2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15		2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16		2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17		2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18		2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19		2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20		2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21		2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22		2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23		2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24		2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25		2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26		2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27		2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28		2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29		2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30		2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40		2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60		1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120		1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞		1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

DENOMINATOR DEGREES OF FREEDOM



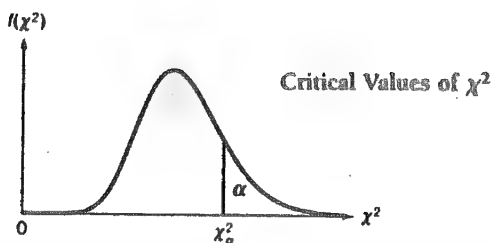
Percentage Points of the F Distribution, $\alpha = .025$

ν_2	NUMERATOR DEGREES OF FREEDOM								
	1	2	3	4	5	6	7	8	9
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11

Source: From M. Merrington and C. M. Thompson, "Tables of Percentage Points of the Inverted Beta (F)-Distribution," *Biometrika*, 1943, 33, 73-88. Reproduced by permission of the Biometrika Trustees.

V ₁	NUMERATOR DEGREES OF FREEDOM									
	10	12	15	20	24	30	40	60	120	∞
1	968.6	976.7	984.9	993.1	997.2	1,001	1,006	1,010	1,014	1,018
2	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

DENOMINATOR DEGREES OF FREEDOM



DEGREES OF FREEDOM	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$
1	.0000393	.0001571	.0009821	.0039321	.0157908
2	.0100251	.0201007	.0506356	.102587	.210720
3	.0717212	.114832	.215795	.351846	.584375
4	.206990	.297110	.484419	.710721	1.063623
5	.411740	.554300	.831211	1.145476	1.61031
6	.675727	.872085	1.237347	1.63539	2.20413
7	.989265	1.239043	1.68987	2.16735	2.83311
8	1.344419	1.646482	2.17973	2.73264	3.48954
9	1.734926	2.087912	2.70039	3.32511	4.16816
10	2.15585	2.55821	3.24697	3.94030	4.86518
11	2.60321	3.05347	3.81575	4.57481	5.57779
12	3.07382	3.57056	4.40379	5.22603	6.30380
13	3.56503	4.10691	5.00874	5.89186	7.04150
14	4.07468	4.66043	5.62872	6.57063	7.78953
15	4.60094	5.22935	6.26214	7.26094	8.54675
16	5.14224	5.81221	6.90766	7.96164	9.31223
17	5.69724	6.40776	7.56418	8.67176	10.0852
18	6.26481	7.01491	8.23075	9.39046	10.8649
19	6.84398	7.63273	8.90655	10.1170	11.6509
20	7.43386	8.26040	9.59083	10.8508	12.4426
21	8.03366	8.89720	10.28293	11.5913	13.2396
22	8.64272	9.54249	10.9823	12.3380	14.0415
23	9.26042	10.19567	11.6885	13.0905	14.8479
24	9.88623	10.8564	12.4011	13.8484	15.6587
25	10.5197	11.5240	13.1197	14.6114	16.4734
26	11.1603	12.1981	13.8439	15.3791	17.2919
27	11.8076	12.8786	14.5733	16.1513	18.1138
28	12.4613	13.5648	15.3079	16.9279	18.9392
29	13.1211	14.2565	16.0471	17.7083	19.7677
30	13.7867	14.9535	16.7908	18.4926	20.5992
40	20.7065	22.1643	24.4331	26.5093	29.0505
50	27.9907	29.7067	32.3574	34.7642	37.6886
60	35.5346	37.4848	40.4817	43.1879	46.4589
70	43.2752	45.4418	48.7576	51.7393	55.3290
80	51.1720	53.5400	57.1532	60.3915	64.2778
90	59.1963	61.7541	65.6466	69.1260	73.2912
100	67.3276	70.0648	74.2219	77.9295	82.3581

Source: From C. M. Thompson, "Tables of the Percentage Points of the χ^2 -Distribution," *Biometrika*, 1941, 32, 188-189. Reproduced by permission of the *Biometrika* Trustees.

DEGREES OF FREEDOM	χ^2_{100}	χ^2_{050}	χ^2_{025}	χ^2_{010}	χ^2_{005}
1	2.70554	3.84146	5.02389	6.63490	7.87944
2	4.60517	5.99147	7.37776	9.21034	10.5966
3	6.25139	7.81473	9.34840	11.3449	12.8381
4	7.77944	9.48773	11.1433	13.2767	14.8602
5	9.23635	11.0705	12.8325	15.0863	16.7496
6	10.6446	12.5916	14.4494	16.8119	18.5476
7	12.0170	14.0671	16.0128	18.4753	20.2777
8	13.3616	15.5073	17.5346	20.0902	21.9550
9	14.6837	16.9190	19.0228	21.6660	23.5893
10	15.9871	18.3070	20.4831	23.2093	25.1882
11	17.2750	19.6751	21.9200	24.7250	26.7569
12	18.5494	21.0261	23.3367	26.2170	28.2995
13	19.8119	22.3621	24.7356	27.6883	29.8194
14	21.0642	23.6848	26.1190	29.1413	31.3193
15	22.3072	24.9958	27.4884	30.5779	32.8013
16	23.5418	26.2962	28.8454	31.9999	34.2672
17	24.7690	27.5871	30.1910	33.4087	35.7185
18	25.9894	28.8693	31.5264	34.8053	37.1564
19	27.2036	30.1435	32.8523	36.1908	38.5822
20	28.4120	31.4104	34.1696	37.5662	39.9968
21	29.6151	32.6705	35.4789	38.9321	41.4010
22	30.8133	33.9244	36.7807	40.2894	42.7956
23	32.0069	35.1725	38.0757	41.6384	44.1813
24	33.1963	36.4151	39.3641	42.9798	45.5585
25	34.3816	37.6525	40.6465	44.3141	46.9278
26	35.5631	38.8852	41.9232	45.6417	48.2899
27	36.7412	40.1133	43.1944	46.9630	49.6449
28	37.9159	41.3372	44.4607	48.2782	50.9933
29	39.0875	42.5569	45.7222	49.5879	52.3356
30	40.2560	43.7729	46.9792	50.8922	53.6720
40	51.8050	55.7585	59.3417	63.6907	66.7659
50	63.1671	67.5048	71.4202	76.1539	79.4900
60	74.3970	79.0819	83.2976	88.3794	91.9517
70	85.5271	90.5312	95.0231	100.425	104.215
80	96.5782	101.879	106.629	112.329	116.321
90	107.565	113.145	118.136	124.116	128.299
100	118.498	124.342	129.561	135.807	140.169

Critical Values of T_L and T_U for the Wilcoxon Rank Sum Test: Independent Samples

Test statistic is the rank sum associated with the smaller sample (if equal sample sizes, either rank sum can be used).

a. $\alpha = .025$ one-tailed; $\alpha = .05$ two-tailed

		3		4		5		6		7		8		9		10	
n ₂	n ₁																
		T _L	T _U	T _L	T _U	T _L	T _U	T _L	T _U	T _L	T _U	T _L	T _U	T _L	T _U	T _L	T _U
3	3	5	16	6	18	6	21	7	23	7	26	8	28	8	31	9	33
4	4	6	18	11	25	12	28	12	32	13	35	14	38	15	41	16	44
5	5	6	21	12	28	18	37	19	41	20	45	21	49	22	53	24	56
6	6	7	23	12	32	19	41	26	52	28	56	29	61	31	65	32	70
7	7	7	26	13	35	20	45	28	56	37	68	39	73	41	78	43	83
8	8	8	28	14	38	21	49	29	61	39	73	49	87	51	93	54	98
9	9	8	31	15	41	22	53	31	65	41	78	51	93	63	108	66	114
10	10	9	33	16	44	24	56	32	70	43	83	54	98	66	114	79	131

b. $\alpha = .05$ one-tailed; $\alpha = .10$ two-tailed

		3		4		5		6		7		8		9		10	
n ₂	n ₁	3		4		5		6		7		8		9		10	
		T _L	T _U	T _L	T _U	T _L	T _U	T _L	T _U	T _L	T _U	T _L	T _U	T _L	T _U	T _L	T _U
3	3	6	15	7	7	7	20	8	22	9	24	9	27	10	29	11	31
4	4	7	17	12	24	13	27	14	30	15	33	16	36	17	39	18	42
5	5	7	20	13	27	19	36	20	40	22	43	24	46	25	50	26	54
6	6	8	22	14	30	20	40	28	50	30	54	32	58	33	63	35	67
7	7	9	24	15	33	22	43	30	54	39	66	41	71	43	76	46	80
8	8	9	27	16	36	24	46	32	58	41	71	52	84	54	90	57	95
9	9	10	29	17	39	25	50	33	63	43	76	54	90	66	105	69	111
10	10	11	31	18	42	26	54	35	67	46	80	57	95	69	111	83	127

Source: From F. Wilcoxon and R. A. Wilcoxon, "Some Rapid Approximate Statistical Procedures," 1964, 20-23. Reproduced with the permission of American Cyanamid Company.

Critical Values of Spearman's Rank Correlation Coefficient

The α values correspond to a one-tailed test of $H_0: \rho_s = 0$. The value should be doubled for two-tailed tests.

n	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$	n	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
5	.900	—	—	—	18	.399	.476	.564	.625
6	.829	.886	.943	—	19	.388	.462	.549	.608
7	.714	.786	.893	—	20	.377	.450	.534	.591
8	.643	.738	.833	.881	21	.368	.438	.521	.578
9	.600	.683	.783	.833	22	.359	.428	.508	.562
10	.564	.648	.745	.794	23	.351	.418	.496	.549
11	.523	.623	.736	.818	24	.343	.409	.485	.537
12	.497	.591	.703	.780	25	.336	.400	.475	.526
13	.475	.566	.673	.745	26	.329	.392	.465	.515
14	.457	.545	.646	.716	27	.323	.385	.456	.505
15	.441	.525	.623	.689	28	.317	.377	.448	.496
16	.425	.507	.601	.666	29	.311	.370	.440	.487
17	.412	.490	.582	.645	30	.305	.364	.432	.478

Source: From E. G. Olds, "Distribution of Sums of Squares of Rank Differences for Small Samples," *Annals of Mathematical Statistics*, 1938, 9. Reproduced with the permission of the Editor, *Annals of Mathematical Statistics*.

TABLE I
Random
numbers

Line number	Column number									
	00-09		10-19		20-29		30-39		40-49	
00	15544	80712	97742	21500	97081	42451	50623	56071	28882	28739
01	01011	21285	04729	39986	73150	31548	30168	76189	56996	19210
02	47435	53308	40718	29050	74858	64517	93573	51058	68501	42723
03	91312	75137	86274	59834	69844	19853	06917	17413	44474	86530
04	12775	08768	80791	16298	22934	09630	98862	39746	64623	32768
05	31466	43761	94872	92230	52367	13205	38634	55882	77518	36252
06	09300	43847	40881	51243	97810	18903	53914	31688	06220	40422
07	73582	13810	57784	72454	68997	72229	30340	08844	53924	89630
08	11092	81392	58189	22697	41063	09451	09789	00637	06450	85990
09	93322	98567	00116	35605	66790	52965	62877	21740	56476	49296
10	80134	12484	67089	08674	70753	90959	45842	59844	45214	36505
11	97888	31797	95037	84400	76041	96668	75920	68482	56855	97417
12	92612	27082	59459	69380	98654	20407	88151	56263	27126	63797
13	72744	45586	43279	44218	83638	05422	00995	70217	78925	39097
14	96256	70653	45285	26293	78305	80252	03625	40159	68760	84716
15	07851	47452	66742	83331	54701	06573	98169	37499	67756	68301
16	25594	41552	96475	56151	02089	33748	65289	89956	89559	33687
17	65358	15155	59374	80940	03411	94656	69440	47156	77115	99463
18	09402	31008	53424	21928	02198	61201	02457	87214	59750	51330
19	97424	90765	01634	37328	41243	33564	17884	94747	93650	77668

APPENDIX A STATISTICAL TABLES

TABLE V (cont.)
Critical values
and approximate
significance levels
for a Wilcoxon
signed-rank test

Sample size n	Significance level, α		Critical value	
	One-tailed	Two-tailed	W_T	W_r
7	0.01	0.02	0	28
	0.025	0.05	2	26
	0.05	0.10	4	24
	0.10	0.20	6	22
8	0.005	0.01	0	36
	0.01	0.02	2	34
	0.025	0.05	4	32
	0.05	0.10	6	30
	0.10	0.20	8	28
9	0.005	0.01	2	43
	0.01	0.02	3	42
	0.025	0.05	6	39
	0.05	0.10	8	37
	0.10	0.20	11	34
10	0.005	0.01	3	52
	0.01	0.02	5	50
	0.025	0.05	8	47
	0.05	0.10	11	44
	0.10	0.20	14	41
11	0.005	0.01	5	61
	0.01	0.02	7	59
	0.025	0.05	11	55
	0.05	0.10	14	52
	0.10	0.20	18	48
12	0.005	0.01	7	71
	0.01	0.02	10	68
	0.025	0.05	14	64
	0.05	0.10	17	61
	0.10	0.20	22	56
13	0.005	0.01	10	81
	0.01	0.02	13	78
	0.025	0.05	17	74
	0.05	0.10	21	70
	0.10	0.20	26	65

TABLE V (cont.)
Critical values
and approximate
significance levels
for a Wilcoxon
signed-rank test

Sample size n	Significance level, α		Critical value	
	One-tailed	Two-tailed	W_1	W_2
14	0.005	0.01	13	92
	0.01	0.02	16	89
	0.025	0.05	21	84
	0.05	0.10	26	79
	0.10	0.20	31	74
15	0.005	0.01	16	104
	0.01	0.02	20	100
	0.025	0.05	25	95
	0.05	0.10	30	90
	0.10	0.20	37	83
16	0.005	0.01	19	117
	0.01	0.02	24	112
	0.025	0.05	30	106
	0.05	0.10	36	100
	0.10	0.20	42	94
17	0.005	0.01	23	130
	0.01	0.02	28	125
	0.025	0.05	35	118
	0.05	0.10	41	112
	0.10	0.20	49	104
18	0.005	0.01	28	143
	0.01	0.02	33	138
	0.025	0.05	40	131
	0.05	0.10	47	124
	0.10	0.20	55	116
19	0.005	0.01	32	158
	0.01	0.02	38	152
	0.025	0.05	46	144
	0.05	0.10	54	136
	0.10	0.20	62	128
20	0.005	0.01	37	173
	0.01	0.02	43	167
	0.025	0.05	52	158
	0.05	0.10	60	150
	0.10	0.20	70	140

- 1- Berenson, L, Mark and David M. Levine "Basic Business statistics Concepts and Applications" 5th ed 1992.
- 2- Box, G. E and Jenkins, G. N. "Time Series Analysis forecasting and Control", 2nd ed, 1977.
- 3- Cochran, W, G, "Sampling Techniques" 3ed ed 1997, New York.
- 4- Conover, W. J. "Practical Nonparametric Statistics" 1971 John Wiley and Sons. Inc.
- 5- Daniel, W. Wayne "Biostatistics. A Foundation for analysis in the health Sciences" 6th ed. 1995 John Wiley and Sons, Inc.
- 6- Fox, John "Regression Diagnostics Thousand Oaks, CA: Sage Publications. Quantitative Applications in the Social Sciences series No 79 provides a through review of methods of of testing the assumptions of regression models 1991.
- 7- Hays, W. L. "Statistics for Social Sciences" 3rd ed 1980 New York. Holt Rinehart and Winston.
- 8- Hogg, N. Robert and Cragg, T, Allen "Introduction to Mathematical Statistics" 5th ed 1995 Prentice - Hall - Inc.
- 9- Larson, Harold, J. "Introduction to Probability" 1995 Addison - Weseley Publishing Company, Inc.
- 10- Larson, Harold, J. "Introduction to Probability Theory and Statistical Inference 3rd ed 1969, 1974, 1982 John Wiley and sons Inc.
- 11- Levine David, M. and Berenson, Mark, and Stephan, David. "Statistics for Managers Using Microsoft excel" 2nd ed 1999, 1997, Prentice-Hall Inc.
- 12- McClave, James, T & Benson, George, P. "Statistics for Business and Economics" 5th ed, 1991 New York.

13- Neter, John and Wasserman, William, "Applied Linear Statistical Models". 1974 Richard D. Win. Inc.

14- Parzen, E. "Modern Probability theory and its application" 1960 New York: Wiley.

15- Waters, Donald "Quntitative Methods for Business" 2nd ed 1993, 1997 Addison Wesley Longman Publishers Ltd.

16- Weiss, Niel, N. "Elementary Statistics" 4th ed 1999, Addison Wesley Longman, Inc.

17- Weiss Neil, N. "Introductory statistics" 5th 1999, Addison Wesley Long man, Inc.

18- أ.د. محمد صبحي أبو صالح "الموجز في الطرق الإحصائية" 2002، دار
اليازوري العلمية للنشر والتوزيع.

19- د. محمود البياتي، "تحليل البيانات الإحصائية باستخدام البرنامج
الإحصائي SPSS" دار الحامد للنشر، 2005م.

20- د. خاشع محمود الراوي، «المدخل إلى الاحصاء»، 1984، جامعة الموصل.